ADVANCED PLATFORMS FOR MULTIPLEXED NUCLEIC ACID RESEARCH

by
Ye Zhang

A dissertation submitted to Johns Hopkins University in conformity with
the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
March, 2019

Abstract

Nucleic acid is the basic genetic material of all life forms on earth. In recent years, nucleic acid technology has assumed an essential role in widespread fields ranging from medicine to agriculture. In the field of molecular diagnostics, nucleic acid analysis has become widely adopted due to their high specificity, sensitivity and their capability for multiplexing. The critical first step in nucleic acid analysis is sample preparation, which involves isolation and purification of DNA/RNA from diverse samples, such as blood and serum. Compared to rapid advances in genomic analysis methods, DNA extraction techniques have remained nearly unchanged over the past 20 years. While spin columns and magnetic microparticles dominate due to their speed and ease-of-use, these methods fragment DNA and are incapable of sufficient DNA quality for the newest long-read sequencing and genome mapping technologies. Moreover, despite all the technological advancements, nucleic acid techniques that allow large-scale multiplexed analysis of polymorphic genetic loci remain needed for theoretical and practical nucleic acid research, such as the study of oncogenic mutations and genetic disease diagnosis. In addition, the cost-effectiveness of infectious disease diagnosis by nucleic acid analysis counts on the multiplex scale in panel testing. So far, solid-phase probe hybridization, PCR and LCR are the three domain assays for locus-specific analysis but none of them can be adapted to the practical needs. Therefore, developing novel and effective strategies for large-scale multiplexed analysis is still a compelling need.

In this thesis, we present advanced platforms for multiplexed analysis of nucleic acids including 1) a simple silica nanomembrane-based method to extract high molecular weight, high purity DNA, 2) a Ratiometric Fluorescence Coding strategy for multiplexed detection of nucleic acids and 3) an easy single-molecule, fluorescence spectroscopic method to measure absolute telomere lengths in a variety of DNA samples.

We start by discussing the existing DNA extraction methods and current strategies for expanding the multiplexed detection capacity of nucleic acid amplification testing. We also introduce the confocal single-molecule spectroscopy and its capability for screening analytes that are not amendable to amplification-based analysis (Chapter 1). Then, we introduce the thermoplastic silica nanomaterial named Nanobind and demonstrate its capability of extracting DNA > 5.7 Mb within 45 mins surpassing any existing methods (Chapter 2). Next, we present and demonstrate our Ratiometric Fluorescence Coding strategy for multiplexed detection of DNA targets from six infectious diseases and the potential for further expanding its multiplexing capability (Chapter 3). Finally, we implement the confocal single-molecule fluorescence spectroscopy to measure the absolute telomere lengths and demonstrate the accurate detection of telomeres as short as 100 bp with high reproducibility and profiling of telomere lengths in human cancer cell lines and primary cells (Chapter 4).

Advisor: Jeff Tza-Huei Wang
Committee members: Dr. Hai-Quan Mao, Dr. Ishan Barman, Dr. Jeff Tza-Huei Wang

Preface

First and foremost, I would like to offer my sincerest gratitude to my advisor, Dr. Tza-Huei Wang, for his consistent encouragement, invaluable advice and unconditional support throughout my Ph.D. Dr. Wang is a great mentor whose guidance has directed me moving forward, inspired me to explore innovative ideas and helped me become an independent researcher. Also, to my thesis committee members, Dr. Hai-Quan Mao and Dr. Ishan Barman, thank you both for providing invaluable feedback throughout the development of this dissertation.

This work would not have been possible without the support from members – past and present – of the BioMEMS lab at Johns Hopkins. I would like to thank Dr. Yi Zhang and Dr. Cyrus Beh for their help in guiding me into decent research directions during the early years of my graduate study. Their work formed the foundation of my research and all the advices they gave and all the skills they taught me have been of excellent value to me. I would like to thank Dr. Kelvin Liu, who have helped me develop new ideas in my experiments and eventually opened a new door to accomplish my first research project in this lab. I would like to thank Dr. Kuangwen Hsieh for improving my writing skills and giving me great advices during my research. I would also like to thank Dr. Thomas Pisanic, who is always pleasant to talk with and offered me valuable help during my graduate study. I am so grateful to be a member of the Wang lab, where every current member and alumni is adorable. Thank you all for your help, your enthusiasm, your kindness and your inspirations throughout all these years, as well as the countless discussions at our lab meetings. Because of each and every one of you, our lab is always an enjoyable place to work in and my graduate study has been energetic and lively. I would also like to thank my research collaborators for providing ideas and expertise that elevated our projects and expanded my thinking. I would also like to thank all my friends, classmates and colleagues for the support and friendship that helped making the journey a lot more fun.

Last but not least, I would like to thank my family. I'd love to thank my beloved parents for their unconditional love, gracious generosity and strong support when their daughter is pursuing her Ph.D. degree 8,000 miles away. Their love, unconditional support and optimistic attitude towards life are always inspiring me whenever I feel frustrated. Their confidence in me makes me feel like I can do anything. I would never have made it without their understanding and support.

I offer my regards and blessings to all of those who supported me in any respect during the completion of the dissertation.

Table of Contents

List of Tables

List of Figures

Chapter 1

1. Multiplexed Nucleic Acid Analysis

1.1 Background

Nucleic acid, one kind of macromolecules present in almost all the cells of most organisms, is one of the most basic substance of life. These macromolecules store, transmit and express genetic information of cells and allow organisms to transfer genetic information from one generation to the next. Based on their different chemical components, nucleic acids can be divided into two types: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). In most cells, DNA plays the major role in the storage, copying and transmission of genetic information, while RNA directs the synthesis of proteins.

Since the landmark discovery in 1953 reporting the double helix structure of DNA [1], there has been a rapid advancement of nucleic acid technology, including recombinant cloning technology [2, 3], polymerase chain reaction (PCR) [4], sequencing [5, 6], CRISPR technology [7, 8], etc. These technologies have been applied and created new commercial opportunities in a wide range of fields. For example, a plethora of medical products, such as insulin and vaccines, have been commercially prepared using recombinant DNA technology for the treatment of many diseases including cancer [9, 10]. Moreover, as the great advances in the development of sequencing techniques, the release of whole genome sequence data for a number of organisms have opened a path for personalized genetic diagnostics and treatment [11, 12]. At the same time, the use of nucleic acid technology has become commonplace in the field of agriculture (e.g. transgenic insect-resistant crops) [13], animal farming (e.g. food safety) [14, 15] and forensic science (e.g. DNA typing).

Among these technological advancement, nucleic acid-based tests/assays have become widely accepted in molecular diagnostics due to their high specificity, sensitivity and their ability to analyse more than one genetic markers simultaneously in each sample [16]. Compare with conventional methods, nucleic acid-based analysis such as PCR, DNA microarrays [17, 18], and automated sequencing techniques can provide more molecular information with easier operation and rapid turnaround time. For example, in the case of pathogen detection, nucleic acid analysis takes much shorter time compared with traditional culture-based methods [19, 20] and enables detection of the presence of pathogen DNA noninvasively from asymptomatic individuals [21]. In general, nucleic acid analysis involves two major steps: 1) sample preparation including the isolation and purification of nucleic acids, and 2) sample detection for the characterization of DNA or RNA of interest (**Figure 1.1**).

1.2 Nucleic Acid Isolation

Isolation of nucleic acids from blood, serum, or plasma, is a critical step in nucleic acid analysis since the quality of the recovered nucleic acid can affect the performance in downstream reactions [22]. Conventionally, liquid phase extraction techniques, such as Phenol/Chloroform precipitation method, are widely used [23-25]. Although this approach yields DNA of better quality, it is laborious, time-consuming and highly operator-dependent. Solid phase extraction techniques are gaining popularity these days [26-28] and they are often the methods of choice when processing a large number of samples. Solid phase nucleic acid extraction via silica adsorption is one of the most common techniques used in modern molecular biology, yet the precise mechanism is not fully understood. It relies on the selective adsorption of nucleic acids to a stationary phase (e.g. porous silica membrane) or a mobile solid phase (e.g. paramagnetic silica microparticles). As a result, the vast majority of commercial nucleic acid extraction kits are performed using either spin columns containing porous silica membranes or silica paramagnetic microparticles [29-31]. These spin columns and micro/nano particles can provide rapid, ease-of-use and generally high purity of DNA extraction, while on the other hand, induce high shearing forces to fragment DNA during flow and mixing, resulting in DNA sample of relatively poor integrity [32, 33].

The ability to obtain high molecular weight DNA with high purity has become increasingly important to ensure accurate downstream genetic analyses such as 3rd generation sequencing. The 3rd generation technologies offer long read lengths surpassing that of 2nd generation and Sanger technologies [34, 35]. For example, Pacific Biosciences' SMRT technology offers read lengths up to 30 kb. The ability to sequence large sections of unfragmented genomic DNA could greatly aid in decoding whole genomes containing long repeat sequences [36]. In gene therapy and DNA vaccines, high binding capacity could be extremely useful in isolating industrial quantities of long plasmids [37-39]. In genetic analysis of telomeres [40] or functional analysis of chromosomal rearrangements such as copy number variation [41], high molecular weight input DNA reduces the chance that fragmentation will occur in a region of interest. Therefore, a method that allows high molecular weight DNA extraction with high yield and purity is under demanding.

1.3. Multiplexed Nucleic Acid Amplification Testing

After sample preparation, the extracted DNA or RNA will be utilized for subsequent characterization and detection via techniques including nucleic acid amplification testing (NAAT), DNA microarray, sequencing, etc. NAAT, which is a technique that involves the amplification and detection of nucleic acids for diagnosis, has emerged as a popular technique for the diagnosis of infectious diseases, such as sextually transmitted diseases [42], and inherited diseases, such as cancer [43, 44], etc. Indeed, NAAT methods, such as real-time PCR [45, 46], have been widely employed for nucleic acid detection and clinical diagnosis due to their high sensitivity, specificity, and rapid turnaround time.

Although NAAT methods have become the cornerstone for molecular diagnosis of diseases, expanding the multiplexed detection capacity of NAAT remains an important objective. Multiplexed NAAT essentially means the ability to simultaneously analyse multiple biomarkers/targets within one sample at the same time, which is particularly useful for clinical diagnosis. For example, the common approach for clinical diagnosis of a patient with nonspecific infection is to prescribe broad spectrum antibiotics and wait for the patient's response. However, the antibiotics may not work at all if the causative pathogen is viral. Then if the physician suspects a particular pathogen and asks for a laboratory test for it, which however, may takes multiple individual tests to rule out the suspected pathogens until the right one is found. This procedure may lead to a long diagnosis time (several hours to even days) and further delay the time for effective treatment. On the other hand, with a multiplexed test capable of screening the patient's sample for multiple suspected bacterial and viral pathogens simultaneously, the physician will be able to identify the pathogen within a single test and expedite the diagnosis process.

Multiplexed NAAT can be accomplished through diverse ways, among which the most commonly used approach is to simultaneously use multiple target-specific probes, each being labelled with a fluorophore of a distinct colour, in a single assay, such as multiplexed real-time PCR with Taq-man probes [47, 48]. However, such one-color-one-target scheme has a limited capacity of multiplexing because of the spectral overlaps between fluorophores, which restrict the number of fluorophores that can be used within a single assay. Several strategies have been developed as alternatives to the one-color-one-target scheme to increase the multiplexing capacity of NAAT methods. For example, Multiplex Ligation-dependent Probe Amplification (MLPA) [49, 50] utilizes a panel of probes with distinct lengths that are ligated after target-specific hybridization and then amplified via PCR to generate amplicons with distinct lengths. The multi-length amplicons in a single reaction – each encoding a specific target – can then be separated and identified using capillary electrophoresis, which however, requires expensive equipment and adds complexity to this assay. Yet another example that obviates electrophoresis is post-PCR melt temperature (T_m) analysis that can distinguish the amplicons generated from specific targets for multiplexed detection [51, 52]. Unfortunately, the number of amplicons/targets that can be identified via T_m is constrained by the temperature resolution and noises from primer dimers, which are hard to avoid in PCR. Therefore, there still remains a need for improving the multiplexing capacity of NAAT methods.

1.4. Single Molecule Detection of Nucleic Acid

Though amplification-based nucleic acid detection has been the keystone for nucleic acid analysis, such enzymatic amplification procedure introduces some challenges for quantitative analysis of nucleic acids. For example, there are errors associated with the amplification procedure that causes nucleic acid sequence misrepresentation, such as base substitutions [53, 54], which may result in false positives in single nucleotide polymorphism (SNP) detection. Moreover, the amplification bias, especially the GC bias in PCR amplification has long been recognized as the major factor that affects the amplification efficiency, causing uneven amplification of different nucleic acid sequences [55, 56]. In addition, there are several classes of analytes that are not amendable to amplification-based analysis including nucleic acid sequences with tandem repeats, such as telomere [57, 58] and trinucleotide repeat in Huntington's disease [59]. Amplification-based methods cannot provide absolute quantification of the tandem repeat length and always suffer from allelic dropout during PCR [60].

Due to these challenges, single molecule detection (SMD) techniques have merged as a powerful approach for nucleic acid analysis, which enables the observation of single nucleic acid molecule with unparalleled detection sensitivity and quantification capability [61, 62]. With the advances in laser-based fluorescence detection in the past half-century, a variety of optical SMD techniques for nucleic acid analysis, which measure individual molecules in solution or on a solid surface, have been developed [63-65]. Optical SMD in solution is usually performed using confocal fluorescence spectroscopy, in which a laser beam is focused into a femtoliter-size detection volume and used to excite single fluorescently-labeled molecules passing through the volume (**Figure 1.2**). The femtoliter detection volume defined by a high-numerical-aperture and a confocal pinhole is capable of greatly reducing the background noise from Raman scattering or undesirable fluorescent impurities compared with bulk detection with large volume (~microliter). Due to its high signal-to-noise ratio, signal from even a single fluorescently-labeled molecule can be effectively detected and distinguished from background. Moreover, a fluorescent burst is detected only when a fluorescently-labeled molecule passing through the small detection volume, thus the total number of molecules in sample can be measured by counting the number of fluorescent bursts. With these unique advantages, optical SMD is suited as an ideal platform for quantitative analysis of low-abundance nucleic acid [66], DNA sizing [67], etc.

However, conventional confocal SMD suffers from two prototypical technical limitations: low mass defection efficiency and quantification artifacts due to the non-uniform Gaussian detection volume. We have developed a new implementation of SMD, termed as cylindrical illumination confocal spectroscopy (CICS) [68], which enables highly quantitative, continuous flow single molecule analysis with high mass detection efficiency and high uniformity. CICS is designed to be a highly sensitive and high throughput platform that can be generically integrated with any microfluidic systems with additional microfluidic components. By employing one-dimensional beam shaping to create a highly uniform sheet-like detection curtain that spans the entire cross-section of a microchannel, CICS achieves near 100% mass efficiency, <10% relative SD in burst heights, and a high signal/noise ratio. It is far less sensitive to thresholding artifacts than traditional SMD and is significantly more accurate at determining both burst rate and burst parameters. This platform has been used to size circulating DNA from serum without amplification [69]. By direct hybridization of unamplified nucleic acids with fluorescently-labeled probes, we expend to implement CICS for quantitative analysis of tandem repeats.

1.5. Summary and Thesis Objectives

In this thesis, we propose to develop an advanced platform for multiplexed analysis of nucleic acid including a simple silica nanomembrane-based method to extract high molecular weight, high purity DNA, a Ratiometric Fluorescence Coding strategy for multiplexed detection of nucleic acid and implementation of the confocal single-molecule spectroscopy for the measurement of absolute telomere lengths in a variety of DNA samples.

In the first chapter, we introduce a novel magnetic thermoplastic silica nanomaterial called Nanobind for high molecular weight DNA extraction. A simple heat-shrinking process is utilized to create a high density of micro- and nano-scale silica lamella that capture DNA using a unique tethering mechanism and harbor it from fragmentation. We compare Nanobind with commercial paramagnetic particles and spin columns, and demonstrate our extraction method enables high yields, high quality and high molecular weight DNA far surpassing that of any existing silica technology. We then demonstrate the high molecular weight and low damage of Nanobind DNA using single molecule, real-time (SMRT) sequencing, which generates exceptional quality long-read libraries and lead to some of the longest reads seen to date on Pacific Biosciences' platform.

Next, we present a new multiplexed detection approach termed Ratiometric Fluorescence Coding, in which we encode each nucleic acid target with a specific ratio between two standard fluorophores. In Ratiometric Fluorescence Coding, we employ the padlock probe chemistry to encode each nucleic acid target with a specific number of binding sites for two probes labelled with different fluorophores. Coupling the padlock probes with either rolling circle amplification (RCA) or hyperbranched rolling circle amplification (HRCA), we transform each nucleic acid target into a specific template that allows hybridization with the fluorescently-labelled probes at pre-designed ratios, thereby achieving multiplexed detection. Furthermore, a droplet-based microfluidic system is coupled with Ratiometric Fluorescence Coding to facilitate high-throughput screening for multiplexed analysis. We demonstrate our approach for multiplexed detection of DNA targets from six infectious diseases and demonstrate its potential for further expanding the multiplexing capability of NAATs.

In the final chapter, we describe an easy single-molecule, fluorescence spectroscopic method named PHAST (hybridization and analysis of single telomere) that permits the profiling of absolute telomere lengths in any DNA sample. We demonstrate the ability of our method to measure the absolute telomere lengths with high reproducibility and minimal calibration. Moreover, we demonstrate its accurate detection of telomeres as short as 100 bp using cloned telomere standards, and have determined the distribution of telomere lengths in human cancer cell lines, primary cells and clinical samples.

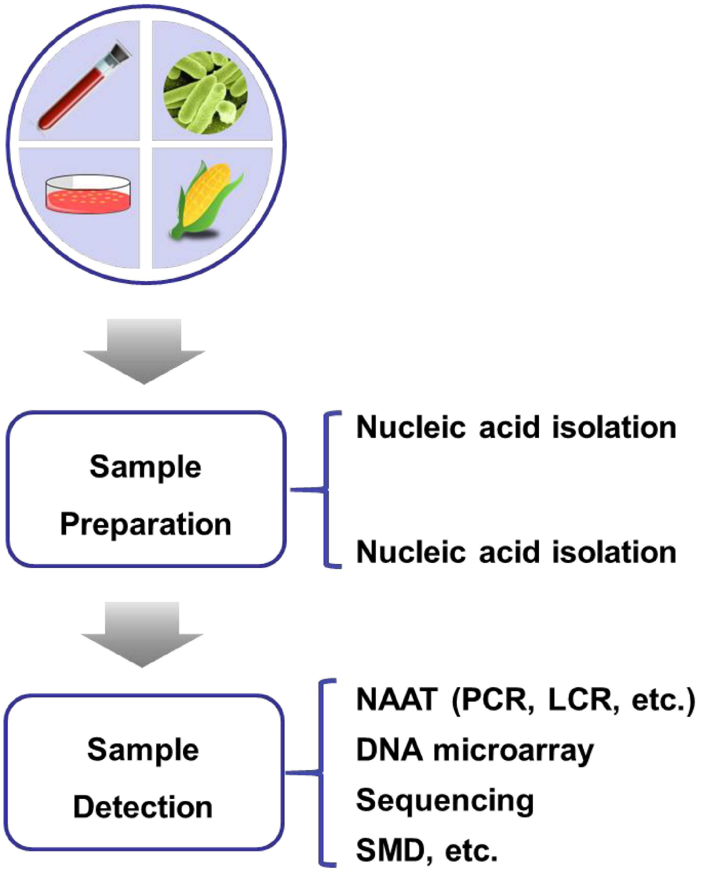

Sample Preparation
Nucleic acid isolation
Nucleic acid isolation
Sample Detection
NAAT (PCR, LCR, etc.)
DNA microarray
Sequencing
SMD, etc.

Figure 1.1. Major steps in nucleic acid analysis workflow

In nucleic acid research, we are interested in studying nucleic acids from all kinds of samples, such as blood, cell lines, bacteria, plants etc. The first critical step is sample preparation, which involves nucleic acid extraction and purification. After sample preparation, the extracted DNA or RNA will be used for detection and analysis via a variety of nucleic acid detection techniques including NAAT (such as PCR), DNA microarray, sequencing, etc.

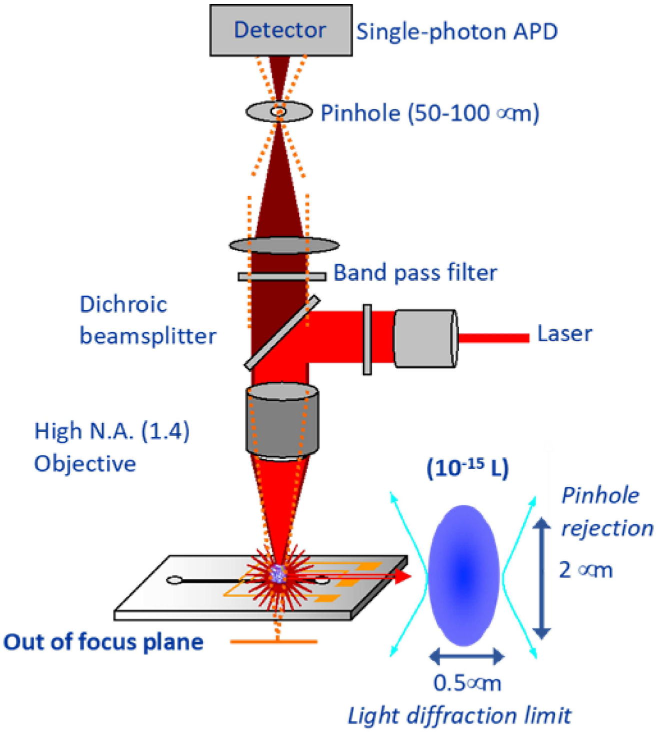

Detector
Single-photon APD
Pinhole (50-100 ∝m)
Band pass filter
Dichroic beamsplitter
Laser
High N.A. (1.4) Objective
(10-15 L)
Pinhole rejection
2 ∝m
Out of focus plane
0.5∝m
Light diffraction limit

Figure 1.2. A schematic view of a confocal SMD spectroscope

Components of a typical SMD spectroscope include a laser source, collimating lens for excitation source, dichroic mirror, objective, sample, emission bandpass filter, focusing lens, confocal pinhole and single photon detector. The detection volume on a confocal SMD spectroscope (blue shaded area; bottom right) is defined by the focused excitation laser beam (bright red) and the collection efficiency of the microscope. The confocal pinhole in the setup (typically 50-100 urn) serves to reject out-of-focus emission from the sample, further defining the detection volume in the vertical axis.

Chapter 2

2. High Molecular Weight DNA Extraction *via* Magnetic Thermoplastic Nanomaterial

2.1 Background

DNA extraction from biological samples, such as blood, serum, or plasma, is a critical step for downstream genetic analysis. The ability to obtain large quantities of high molecular weight (HMW) and high-quality DNA is of rising importance as genomic analysis tools become increasingly sophisticated and the implications of long-range genomic information are realized. In DNA sequencing, 3rd generation technologies, such as Pacific Biosciences' SMRT technology and Oxford Nanopore's MinION technology, offer long read lengths greatly surpassing Sanger sequencing that facilitate de novo sequence assembly and the deciphering of complex repeat sequences [34, 35, 41, 70]. Barcoding technologies such as 10X Genomics' GemCode offer the potential to link myriad short reads in Mb-sized long-reads [71]. Recent genome mapping methods such as BioNano Genomics' Irys technology and OpGen's Argus technology require DNA that is hundreds of kb to Mb in length in order to characterize chromosomal rearrangements and long-range genetic information such as copy number variation and inversions [72, 73].

Despite rapid advances in such genomic tools, few advances have been made in DNA extraction since silica extraction technology was first introduced [74, 75]. To achieve high DNA binding capacities (i.e. extracted DNA per unit area) and extraction yields (i.e. total amount of extracted DNA), these methods rely on the adsorption of DNA to micro- and nanostructured materials with high surface area to volume ratios, such as magnetic microparticles [29, 76-79] or spin columns containing porous materials [78, 80]. These methods have enjoyed widespread adoption due to their speed, ease-of-use, and generally high performance (i.e. high purity and efficiency). However, the shear forces imposed by the existing silica extraction methods tend to fragment DNA such that tedious precipitation methods (e.g. phenol-chloroform or salting out) and multi-day gel plug extractions [81] remain the only option when the highest DNA quality is required.

In this chapter we introduce a conceptually different approach to silica-based high molecular weight DNA extraction by developing a novel magnetic silica nanomaterial called Nanobind. This non-porous, planar nanomembrane contains a hierarchical layering of microscale folds and nanoscale silica lamella. In contrast to microparticles and spin columns where nucleic acids can be sheared by particle mixing or by flow through porous matrices, the high surface area of silica lamella on Nanobind protect nucleic acid from shear force-induced fragmentation by harboring it in a condensed, tentacle conformation that enables the extraction of vast amounts of ultra-high molecular weight (UHMW) DNA. This gentle process ensures rapid isolation (<45 minutes) of high quality, megabase (Mb) DNA, containing minimal damage such as nicks and abasic sites, which can be used to generate high quality, single molecule sequencing libraries capable of superior read lengths. Furthermore, Nanobind is easily magnetized to facilitate sample processing. Herein, we explore Nanobind's unique binding mechanism, the fabrication parameters affecting silica lamella surface topography, and characterize the DNA extraction performance. We also perform a thorough comparison against the three most commonly used DNA extraction methods and conclude by using Nanobind extracted DNA for long-read sequencing using Pacific Biosciences single molecule, real-time (SMRT) sequencing technology.

2.2 Experimental Details

2.2.1 Nanobind Fabrication

Non-magnetic Nanobind was fabricated by depositing silicon dioxide onto both sides of a 20 μm thick polyolefin film (Uline, Pleasant Prairie, WI) using electron beam evaporation. Then the oxide-deposited polyolefin film was baked in a convection oven at 165 °C for 3 minutes to heat shrink and induce surface wrinkling. The film shrunk isotropically by 78% in each linear direction. A 6 mm disk was then created using a metal punch. Magnetic Nanobind was fabricated by depositing 100 nm of silicon dioxide on side 1 of a 20 μm thick polyolefin film (Uline, Pleasant Prairie, WI) and then depositing a layer stack containing 20 nm silicon dioxide, 30 nm iron, and 20 nm silicon dioxide. Heat shrinking and die punching were performed identically to non-magnetic Nanobind.

2.2.2 BET Measurements

The BET surface area was measured using a Micromeritics ASAP 2020 physisorption analyzer (Micromeritics, Norcross, GA). Sample tubes containing 85 and 65 disks of the 100 nm oxide and 200 nm oxide Nanobind substrates, respectively, were degassed for 10 hrs at 90 °C. Krypton adsorption isotherms were then measured at 77 K and relative pressures ranging from 0.03 to 0.5. The BET surface area was calculated from the isotherms using the standard method [82].

2.2.3 Nanobind DNA Extraction

To eliminate differences from extraction chemistry, Nanobind DNA extraction was performed using the same buffer system as microparticle and spin column extractions. All extractions were performed using a single 6 mm Nanobind disk in a 1.5 mL microcentrifuge tube. Unless otherwise indicated, non-magnetic Nanobind was used. λ DNA extraction was performed by adding 200 μL of Buffer AL (Qiagen, Valencia, CA) to 10 μg of λ DNA (New England Biolabs, Ipswich, MA) and mixing well. Cultured cell extraction was performed by adding 50 μL of 10 mg/μL proteinase K (Life Technologies, Carlsbad, CA) and 200 μL of Buffer AL to trypsinized cells that had been resuspended in PBS and incubating at 55 °C for between 10 minutes to 1 hr. For both sample types, a 6 mm Nanobind disk and 300 µL of isopropanol was added, and the mixture was rotated at room temperature for 10-30 minutes. This allowed DNA to bind to the Nanobind surface. After binding, the liquid in the tube was pipetted out and discarded. Then 700 µL of buffer AW1 (Qiagen, Valencia, CA) was added to wash the Nanobind substrate. After pipetting out and discarding buffer AW1, 500 µL of buffer AW2 (Qiagen, Valencia, CA) was added to wash the Nanobind substrate a 2nd time. The AW2 washing step was repeated once more, and all the liquid was removed leaving only the Nanobind substrate and bound DNA in the tube. To ensure complete evaporation of any residual liquid, the tube was heated at 55 °C for 1-2 minutes with the cap open. Afterwards, 100 µL of DI water was added into the microcentrifuge tube, and the tube was incubated at 70°C for 10-20 minutes to elute the DNA. Finally, the eluent was transferred into a new tube.

2.2.4 UHMW Nanobind DNA Extraction

Extractions were performed identically to above except buffer AL, buffer AW1, and buffer AW2 were replaced by custom formulations developed by Circulomics called CL1, PW, and SW, respectively. Wide bore pipette tips and gentle pipetting were also used to minimize DNA shear due to fluid handling.

2.2.5 Microparticle DNA Extraction

Microparticle DNA extraction was performed with the same buffers as Nanobind using the Qiagen BioSprint 96 DNA protocol. First, lysis was performed identically to the Nanobind extraction process above. Then 30 µL of MagAttract Suspension G magnetic particles (Qiagen, Valencia, CA) was added to the sample. The tube was incubated at room temperature for 15 minutes for binding and then placed on a magnetic holder to pull the particles aside such that the liquid could be removed. After the tube was removed from the magnetic holder, 500 µL of buffer AW1 was added and mixed with the microparticles. Then the supernatant was discarded by placing the tube on the magnetic holder again and pipetting out. Next, the magnetic particles were washed two more times with 500 µL of buffer AW2, and the tube containing the magnetic particles was dried at 55 °C for 3 minutes with an open cap. After all the liquid in tube had evaporated, 100 µL of water was added, and the mixture was incubated at 70 °C for 20 minutes. Finally, the tube was put onto the magnetic holder to facilitate the collection of the eluent.

2.2.6 Nanobind DNA Preparation for SEM

After binding genomic DNA onto the Nanobind surface, the Nanobind disk with DNA was fixed with buffer (3% formaldehyde, 1.5% Glutanraldehyde in 0.1 M NaCacodylate, 5mM Ca2+, 2.5% Sucrose, pH 7.4) at room temperature for 10 minutes and then washed in 0.1 M Cas/2.5% sucrose pH 7.4 three times. The disk with DNA was then post-fixed with Palade's 1% OsO_4 for 15 minutes at room temperature. and rinsed in Kellenberger UA (20% Acetate-veronal stock, 28% 0.1 M HCl, 5 g/L Uranyl Acetate and ddH_2O, adjusting pH to 6 with UA) for 1 min at room temperature. After a quick rinsing in ddH_2O, the disk was dehydrated with graded series of cold ethanol (50%, 70%, 95%, 100%) and then washed in fresh 100% ethanol at room temperature. Finally the Nanobind disk with DNA was air dried at room temperature for 30 mins prior to SEM scanning.

2.2.7 Spin Column DNA Extraction

Spin column DNA extraction was performed using the QIAamp DNA Mini kit (Qiagen, Valencia, CA) according to the manufacturer's protocol.

2.2.8 Phenol-Chloroform DNA Extraction

Phenol-chloroform DNA extraction began by adding 300 µL of lysis buffer (50mM Tris, 50mM EDTA and 2% SDS) and 50 µL of 10 mg/mL proteinase K (Life Technologies, Carlsbad, CA) to the cultured cells. After incubation at 55°C overnight, the digested sample was transferred into a MaXtract High Density phase lock gel tube (Qiagen, Valencia, CA) and mixed with 350 µL of 25:24:1 phenol:chloroform:isoamyl alcohol by shaking until the mixture turned milky white. The tube was then centrifuged at 15,000 x g for 5 minutes, and the upper aqueous phase containing the DNA was pipetted out of the phase lock gel tube into a fresh microcentrifuge tube. Next, 35 µL of pH 7.5 ammonium acetate and 700 µL of 100% ethanol was added, and the mixture was incubated at -20°C overnight to precipitate the DNA. Then, the sample was centrifuged at 13,000 rpm for 10 minutes. After discarding the liquid waste, the DNA pellet was washed in 300 µL of 75% ethanol, and the tube was centrifuged a 2nd time at 13,000 rpm for 5 minutes. All the ethanol was discarded after the washing step, leaving only the DNA pellet in the tube. The tube was incubated at 60°C for around 1 to 2 minutes with the cap open to evaporate any remaining ethanol. Finally, 100 µL of DI water was added into the tube, and the tube was incubated at 50°C for 20 minutes to rehydrate the DNA pellet.

2.2.9 DNA Quantification and UV Absorbance

Quant-iT PicoGreen (Invitrogen, Eugene, OR) was used to quantify the isolated DNA. The PicoGreen dye was first diluted 1:200 in 1X TE buffer (Invitrogen, Eugene, OR) while the isolated DNA sample was diluted 1:10 or 1:20 in 1X TE buffer. 50 µL of the diluted DNA sample was mixed with 50 µL of the diluted PicoGreen dye solution, and the mixture was incubated in the dark at room temperature for around 5 minutes. The fluorescence was then measured on a NanoDrop 3300 fluorospectrometer (Thermo Scientific, Wilmington, DE). The sample DNA concentrations were determined by comparing the fluorescent signal against a λ DNA serial dilution standard curve. The purity of isolated DNA was evaluated by UV absorbance using a NanoDrop 1000 UV-Vis spectrophotometer (Thermo Scientific, Wilmington, DE).

2.2.10 Agarose Gel Electrophoresis

The size of the extracted DNA was evaluated using a 0.5% agarose gel in 1X TAE buffer and running at 8V/cm for 90 minutes. The gel was stained using GelStar nucleic acid gel stain (Lonza, Allendale, NJ) and imaged with a Typhoon 9400 variable mode imager (GE Healthcare, Piscataway, NJ).

2.2.11 Pulsed Field Gel Electrophoresis

Pulsed field gel electrophoresis was performed using a 1% Megabase certified agarose gel (Bio-Rad, Hercules, CA) in 0.5X TBE buffer and running at 14° C, 120° angle, 6V/cm, and 35s - 90s switch ramping for 22 hours on a CHEF DR III system (Bio-Rad, Hercules, CA). The gel was stained using SYBR Gold (ThermoFisher, Eugene, OR) and imaged with a Typhoon 9400 variable mode imager (GE Healthcare, Piscataway, NJ). The reference DNA sample was a lambda ladder (CHEF DNA Size Standard #170-3635, Bio-Rad, Hercules, CA). For analysis of UHMW DNA, PFGE was performed using a 1% Megabase certified agarose gel (Bio-Rad, Hercules, CA) in 1X TAE buffer and running at 14° C, 106° angle, 2V/cm, and 1800s switch time for 68 hours and an S. pombe ladder (Bio-Rad, Hercules, CA).

2.2.12 Single Molecule Free Solution Hydrodynamic Separation

All DNA samples were stained separately at 5 ng/µL in 10 mM EACA, 10 mM Bis-Tris buffer with 1 µM TOTO-3 and 500 pM of a reference dye (Alexa647) for at least 1 hour in the dark. The reference DNA sample was a 5 kb ladder (CHEF DNA Size Marker #170-3624, Bio-Rad), which was stained under the same conditions as the extracted genomic DNA samples. All separations were performed in the same 600 cm effective length, 10 µm ID fused silica microcapillary (Polymicro Technologies, Phoenix, AZ). The loading and elution buffers consisted of 10 mM EACA, 10 mM Bis-Tris, and 25 mM NaCl. Sample plugs were loaded into the capillary using a 10 second, 50 psi injection. Separation was performed at 225 psi. Capillaries were flushed with clean elution buffer between samples.

The raw APD counts were collected in 0.1 ms bins. The raw APD data were then normalized for APD dead time and quantum efficiency to obtain fluorescence photon counts. The normalized fluorescence traces could then be analyzed via single molecule counting or bulk fluorescence analysis.19 For bulk fluorescence analysis, the fluorescence traces were re-binned into 6 second bins and smoothed in OriginPro (OriginLab, Northampton, MA) using the Savitzky-Golay method with a window size of 5 points and a polynomial of order 2. Background fluorescence was calculated as the average of one minute of data immediately prior to the beginning of each DNA separation, and was subtracted from the total fluorescence before plotting. The retention times across all the samples were normalized against the reference dye retention time.

2.2.13 qPCR Analysis

qPCR was performed using 20 mM Tris-HCl, 10 mM (NH4)2SO4, 10 mM KCl, 2 mM MgSO4, 0.1% Triton® X-100, 10 mM dNTPs, 200 nM PCR primers, and 1.25 units per 50 µL of Taq polymerase (New England Biolabs, Ipswich, MA) in a 25 µL reaction volume. PCR primers were synthesized (Integrated DNA Technologies, Coralville, IA) to amplify a 148 bp region of the human GAPDH gene (Fwd Primer: 5’ - GAA GGT CGG AGT CAA CGG ATT T / Rev Primer: 5’-ATG GGT GGA ATC ATA TTG GAA C). qPCR was performed in a CFX96 Touch (Bio-Rad, Hercules, CA) using EvaGreen (Bio-Rad, Hercules, CA) with the following thermal cycling parameters: 30 sec at 95°C, 35 cycles of 30 sec at 95°C, 1 minutes at 60°C and 1 minutes at 72°C, and a final extension of 10 minutes at 72°C. Threshold cycle (Ct) was determined using Bio-Rad's CFX Manager software. Each DNA extraction sample was used to generate 10-fold serial dilutions and analyzed in triplicate.

2.2.14 SMRTbell Library Preparation and Sequencing

DNA library preparation and sequencing was performed at Pacific Biosciences according to the standard >30kb SMRTbell library preparation guide available on PacBio SampleNet. In short, 5–20 µg of high-quality, genomic DNA for each sample was verified via Qubit to quantify the amount of dsDNA present. After quantification, each sample was diluted to ~250 ng/µl in PacBio elution buffer. For several of the samples, DNA amounts were quantified after shearing, as the high molecular weights of these samples made quantification difficult. Samples were subject to needle shearing by passing the sample through a 26G needle five to forty times (26G needles from SAI Infusion Technologies, 1 mL Luer-Lok tip syringes from Becton Dickinson) The sheared genomic DNA was then repurified using a 0.45× AMPure PB purification step to ensure removal of very short fragments (<1 Kb) and biological contaminants before proceeding with the enzymatic steps in the library preparation process.

Following library construction, the DNA templates were size selected for a lower cutoff of 30 kb using a BluePippin instrument (Sage Science, MA, USA). Libraries were subject to a post size-selection damage repair step, and the sizes of the resultant libraries were confirmed using pulsed field gel electrophoresis (CHEF Mapper XA, Bio-Rad). Sequencing was performed on the PacBio RS II using P6 polymerase binding and C4 sequencing kits. Samples were MagBead loaded with on chip concentrations ranging from 25 to 50 pM with 360 min acquisition times.

2.3 Silica Lamella Fabrication and Characterization

The high surface area of nanoscale lamella structures on Nanobind are fabricated via a simple and inexpensive 2-step process comprising evaporation of silicon dioxide onto both sides of polyolefin (PO) film and heat-shrinking in an oven (**Figure 2.1A**). Film stresses generated by the heat shrinking process lead to a hierarchical layering of microscale polyolefin folds and nanoscale silica lamella that can act as a high surface area, low shear substrate for DNA extraction. Optionally, a magnetic sandwich layer, which imparts a dark gray hue to the white substrate, may be deposited inline to facilitate magnetic manipulation. In this case, silicon dioxide is used to sandwich an iron layer such that DNA adsorption can continue to occur on both sides of the

Nanobind substrate while also passivating the metallic iron layer from unwanted buffer interactions. DNA isolation is performed using a standard bind, wash, and elute protocol that parallels spin column and magnetic microparticle extraction (**Figure 2.1B**) and can be completed in less than 45 minutes using either non-magnetic or magnetic processing. Magnetic processing can further be performed in either manual or automated format using standard robotic instruments (**Figure 2.2**).

Nanobind DNA is captured via a novel tentacle binding mechanism in which a macromolecular entanglement of DNA is tethered to the silica substrate (i.e. the white disk) at a single end and extends deep into solution in a manner akin to jellyfish tentacles **(Figure 2.1C)**. The low shear format prevents fragmentation and allows Nanobind to act as a seed upon which high molecular weight DNA tentacles grow. As cellular input increases, the DNA assembles into sheet- and sack-like structures, enabling a small area of substrate to capture vast amounts of DNA. SEM images (**Figure 2.3**) show that HMW genomic DNA assembles into sheet-like structures and that tether across silica lamellae surface.

High surface area is created by the hierarchical layering of nanoscale silica lamella and microscale polyolefin folds. As seen in **Figure 2.4**, the surface topography can be fine-tuned by varying the amount of silicon dioxide that is deposited prior to heat shrinking. With 2 nm of silicon dioxide, the oxide layer largely conforms to the microfolds of the underlying polyolefin film. With 20 nm and 50 nm of silicon dioxide, nanoscale silica wrinkles and lamella begin to develop on top of the microfolds as the film stresses begin to overwhelm the adhesion between the oxide and underlying polyolefin. With 100 nm, 150 nm, and 200 nm of oxide, a high density of silica lamella begin to form and progressively increase in size and number as film stresses cause the silicon dioxide to delaminate and fracture. In addition, rather than fully detach and release, the silica lamella remain stably anchored in place by a combination of mechanical interlocking, residual adhesive portions, and lamina-lamina adhesion. Proper design of the layer structures further ensures that addition of the optional magnetic sandwich does not affect silica lamella surface topography

(**Figure 2.4G**). SEM analysis also shows that the lamella remain stably anchored before and after DNA extraction with no observable degradation (**Figure 2.5**).

This hierarchical layering of features provides a substantial increase in surface area over a planar substrate. Using Brunauer-Emmett-Teller (BET) gas adsorption, we measured the surface area of the 100 nm and 200 nm oxide Nanobind substrates. Each 6 mm diameter disk resulted in BET surface areas of 9.33 cm^2 and 9.74 cm^2, respectively, a 35-fold increase over the planar substrate (**Figure 2.6**).

2.4 Nanobind DNA Binding Mechanism

To investigate the effects of silica lamella topography on DNA binding, we performed extractions using the 0 nm oxide (i.e. no oxide control), 20 nm oxide (i.e. low density lamella), 100 nm oxide (i.e. high density lamella), 200 nm oxide (i.e. high density lamella), and unshrunk 100 nm oxide (i.e. flat oxide control) Nanobind substrates. As expected based on SEM and BET analysis, DNA extraction efficiency progressively grew with oxide thickness and the expected surface area (**Figure 2.7A**). Both the 100 nm and 200 nm oxide Nanobind substrates had high efficiency and recovered nearly all of the input DNA. In contrast, the unshrunk 100 nm oxide Nanobind substrate recovered only moderate amounts of DNA while the 0 nm oxide Nanobind substrate recovered very little DNA, demonstrating that a high density of silica lamella is critical for high extraction yields. The non-magnetic and magnetic Nanobind substrates can be used interchangeably with no noticeable difference in performance (**Figure 2.8**).

Next, we attempted to measure Nanobind binding capacity by performing extractions on cultured cells. As seen in Figure 3b, extraction yield increased linearly across the entire range of cell inputs. With $25x10^6$ MCF-7 cells, the 6 mm Nanobind extracted over 175 µg of DNA, representing 100% efficiency and a binding capacity of 188,000 $\mu g/m^2$. This binding capacity is 100-1,000,000-times greater than previously reported values for silica microparticles (<.03 $\mu g/m^2$

to 2,100 µg/m^2) [29, 76-80].The true Nanobind binding capacity likely exceeds this as binding saturation had not yet been reached. It is noteworthy that with 25x10^6 cells input, over 600 µg of RNA was also co-purified, further illustrating the high binding capacity.

A direct comparison against commercial spin columns and magnetic microparticles showed that with moderate cell input (**Figure 2.9A**), Nanobind obtained twice the amount of DNA, but with a high cell input (**Figure 2.9B**), Nanobind yielded 10 to 30-times the amount of DNA. We predict this difference would continue to grow with further increases in cell input. The 1.5 mL format spin column and microparticle processes used herein are capable of high efficiency when extracting 10 to 20 µg of DNA but decrease in efficiency at higher inputs as the binding surfaces become saturated. Using Nanobind, over 200 µg of DNA could be extracted and eluted into a volume as small as 50 µL, resulting in DNA concentrations as high as 4 µg/µL. Nanobind also demonstrated high repeatability with extraction-to-extraction CV of only 12%.

Despite the increased extraction yield, the Nanobind surface area (9.33 cm^2) was actually less than that typically used in microparticle extractions (45-275 cm^2) [76]. Thus, we postulated that the extreme binding capacity was actually a result of the unique tentacle binding mechanism. Thus, large genomic DNA fragments, which more easily form macromolecular tentacles than short DNA fragments, should result in higher extraction yields. To test this, we measured Nanobind binding capacity using λ DNA that had been heat fragmented from 48 kb progressively down to 400 bp (**Figure 2.10**). As seen in Figure 2.7c and Figure 2.10, Nanobind is exceptionally efficient at capturing large DNA. Full length λ DNA exhibited a binding capacity 80-times greater than 400 bp DNA. The true difference is potentially even greater as the full length λ DNA had not reached binding saturation at 100 µg of input. This dramatic increase in large DNA binding capacity can be attributed to the decrease in surface area necessary to tether a few pieces of large DNA as opposed to many pieces of small DNA.

As evident in **Figure 2.7C** and **Figure 2.4G**, the bound genomic DNA does not lie in a strictly prone conformation. Previous studies have modeled the physics of DNA adsorption [29, 76, 83, 84] while attempting to untangle complex electrostatic, salt-bridging, and entropic binding effects [29, 76, 78]. The simplest models assume that DNA lies prone against the silica substrate, bound by multiple contact points along the length of the molecule.[5] This mechanism implies that binding capacity is limited predominantly by substrate surface area and DNA packing density, with a maximum surface coverage of 20-30% [29, 79]. Studies have further proposed that binding conformation can be affected by parameters such as buffer conditions, surface properties, packing density, and, potentially, DNA length [76]. Through the use of the low shear, planar, non-porous Nanobind, we are able to bias the DNA binding away from the typical prone conformation towards a tethered conformation that is particularly efficient for high molecular weight DNA extraction. It is possible that due to this unique mechanism binding saturation may never reasonably occur.

2.5 DNA Extraction Comparison

We then compared Nanobind (NB) against phenol-chloroform (PC), spin column, and magnetic microparticles by extracting genomic DNA from $3x10^6$ cultured MCF-7 cells using all four methods. First, we compared the length of the extracted DNA using single molecule free solution hydrodynamic separation (SML-FSHS) [67, 85]. SML-FSHS uses single molecule sensitivity to enable accurate determination of large DNA size and concurrent detection of minute amounts of DNA fragmentation. As shown in Figure 3d**,** Nanobind and phenol-chloroform DNA had high molecular weights with length centered around 100 kb. Nearly all of the DNA was greater than 50 kb with substantial portions greater than 100 kb. Magnetic particle DNA showed increased fragmentation with very little DNA >90 kb while spin column DNA displayed obvious shearing and appeared as a broad smear. Agarose gel electrophoresis confirmed these trends (**Figure 2.11**) while pulsed field gel electrophoresis further showed that Nanobind extracted DNA up to 300 kb with an average length around 150 kb (**Figure 2.7E**). Under UV absorbance analysis, all four systems showed high purity with 260/280 ratios between 1.99 to 2.04 and 260/230 ratios between 2.10 and 2.25 (**Figure 2.7F** and **Figure 2.12**). When qPCR was used to analyze 10-fold serial dilutions of the previously extracted DNA, all 4 samples exhibited the expected log-linear relation between DNA input and threshold cycle C_t (**Figure 2.7G** and **Figure 2.13)**. At all dilutions, Nanobind and phenol-chloroform extracted DNA resulted in lower C_t values than the microparticle or column extracted DNA. As the same number of cells was input into each extraction, the lower C_t value is likely a combination of factors such as higher extraction yield, lower amounts of carryover PCR inhibitors (e.g. residual organic solvents, salts, etc), and lower DNA damage.

2.6 Long-Read Single Molecule Sequencing

We selected Nanobind and phenol-chloroform extracted DNA to create long-insert SMRTbell libraries for single molecule sequencing as these samples exhibited the highest molecular weights. For Nanobind DNA, an on-substrate RNase digestion and wash were added to remove RNA contamination and obtain pure genomic DNA. Both samples were capable of supporting libraries with a 30 kb lower cutoff, the maximum cutoff possible using the current Blue Pippin size-selection method (**Figure 2.14**). The mean maximum subread lengths per zero-mode waveguide (ZMW) produced by the phenol-chloroform and Nanobind libraries were 20.6 kb and 20.1 kb, respectively. The subread N50 values, based on the longest reads obtained per ZMW, were 29.4 kb for phenol-chloroform (**Figure 2.15A**) and 28.0 kb for Nanobind (**Figure 2.15B**), indicating that 50% of the sequenced bases are in subreads greater than 29.4 and 28.0 kb, respectively. For both samples, approximately 10% of subreads were in excess of 40 kb, and the longest reads were near 70 kb. Consistent with these outstanding subread length distributions, the frequency of damage-related sequencing terminations in these libraries is estimated to be less than one in 500 kb, indicating five- to ten-fold less damage than that typically observed in libraries of this size.

To further test the capabilities of Nanobind, we developed an UHMW process using a custom buffer set and protocol to further reduce shear. Using this process, extracted DNA size rose from 300 kb up to greater than 5.7 Mb (**Figure 2.15C** and **Figure 2.16**). To our knowledge, this is the largest DNA ever extracted using a standard, silica-based bind, wash, and elute protocol. The UHMW DNA was then used to generate a SMRTbell library with 30 kb cutoff as before and subjected to single molecule sequencing. The resultant library produced a mean maximum subread length of 16.7 kb, a subread N50 value of 24.6 kb, and maximum subread length of 73.7 kb (**Figure 2.15D, E**). While these sequencing metrics are not as outstanding as the first set of sequencing data, they are consistent with libraries of the highest quality, and the differences likely reflect variability in template preparation and chip loading.

2.7 Conclusion

In this chapter, we have demonstrated Nanobind for extracting DNA with ultra high MW exceeding Mb in length, low damage, and high purity. This is evidenced by both its high performance in qPCR and in long-read single molecule sequencing. Using Nanobind DNA, we prepared the highest quality long-insert libraries and generated extremely long single molecule reads, with 50% of the sequenced bases contained in subreads greater than 28 kb, and maximum reads in excess of 70 kb. These read length gains have important implications for genome assembly, as there is a tradeoff between read length, coverage, and expected contig length in a genome assembly. These gains can be roughly predicted by the Lander-Waterman model [86], a widely applied model from the Sanger sequencing era. This model predicts that, under ideal conditions in an assembly in which all overlaps have been detected perfectly, expected contig length for a given coverage level should scale linearly with increasing read lengths (**Figure 2.17A**). Thus, increasing read length from 15 to 30 kb is predicted to increase the average contig length by two-fold. Looking at it another way, the amount of coverage needed to achieve a comparable expected contig length decreases with increasing read length (**Figure 2.17B**). In practice, the simplifying assumptions made by this model, particularly the assumption that the genome is free of repeats, mean that this model effectively gives an upper bound on what is achievable. In reality, the relationship between read length, coverage, and contig size is dependent on the number and length of repeats in a given genome. Longer reads additionally allow for sequencing through longer repeat regions, potentially allowing access to previously inaccessible regions of a genome and improving genome assembly.

These data suggest that Nanobind is an excellent alternative to phenol-chloroform and gel plug extraction when the highest quality genomic DNA is required. Whereas these methods can take 1-3 days and require substantial manual skill, the equivalent Nanobind extraction can be performed in less than 45 minutes. Nanobind offers a unique combination of high binding capacity and high molecular weight DNA that cannot be achieved with any existing solid phase technology. The high binding capacity enables high extraction yield while the short processing time and non-porous format render it easily incorporated into small volume, microfluidic, and automated extraction systems at µL to mL scales. Additionally, the ability to re-suspend into small volumes is useful in obtaining high concentration DNA. This suggests that Nanobind can also be used as a plug-and-play substitute for spin columns and microparticles in diverse research, clinical, and production applications beyond high molecular weight. Furthermore, by engineering Nanobind to contain a magnetic sandwich on one side, we are able to process samples manually using a magnetic rack or automated using a robotic instrument.

Film stresses in polymer substrates have been previously used to create surfaces with controlled wrinkle structures [87] while heat shrinking of thermoplastic films has been used to create precision microfluidic channels [88] and micropatterned surfaces [89]. Alternatively, stresses within layer-by-layer deposition processes have been used to create superhydrophobic, hierarchical metal nanoflake surfaces [90]. In this chapter, we integrate these principles to create a novel hierarchical silica lamella surface. The underlying polyolefin film costs pennies per square foot, and the fabrication can be easily scaled using industrial roll-to-roll processes to create a nanostructured DNA extraction substrate that is exceptionally inexpensive to produce. The extraction chemistry incorporated herein used off-the-shelf buffers and obtained a mixture of both DNA and RNA but high efficiency extractions have also been performed using our own custom buffers. For example, in the UHMW process, custom buffers and protocols were able to push extracted DNA length north of 5.7 Mb, while retaining a short 45 minute process time. Future work will integrate Nanobind into other molecular analysis workflows including genome mapping, methylation analysis, and microRNA profiling from a broad array of samples including bacterial and plant cells, tissues, body fluids, and formalin-fixed paraffin embedded (FFPE) samples.

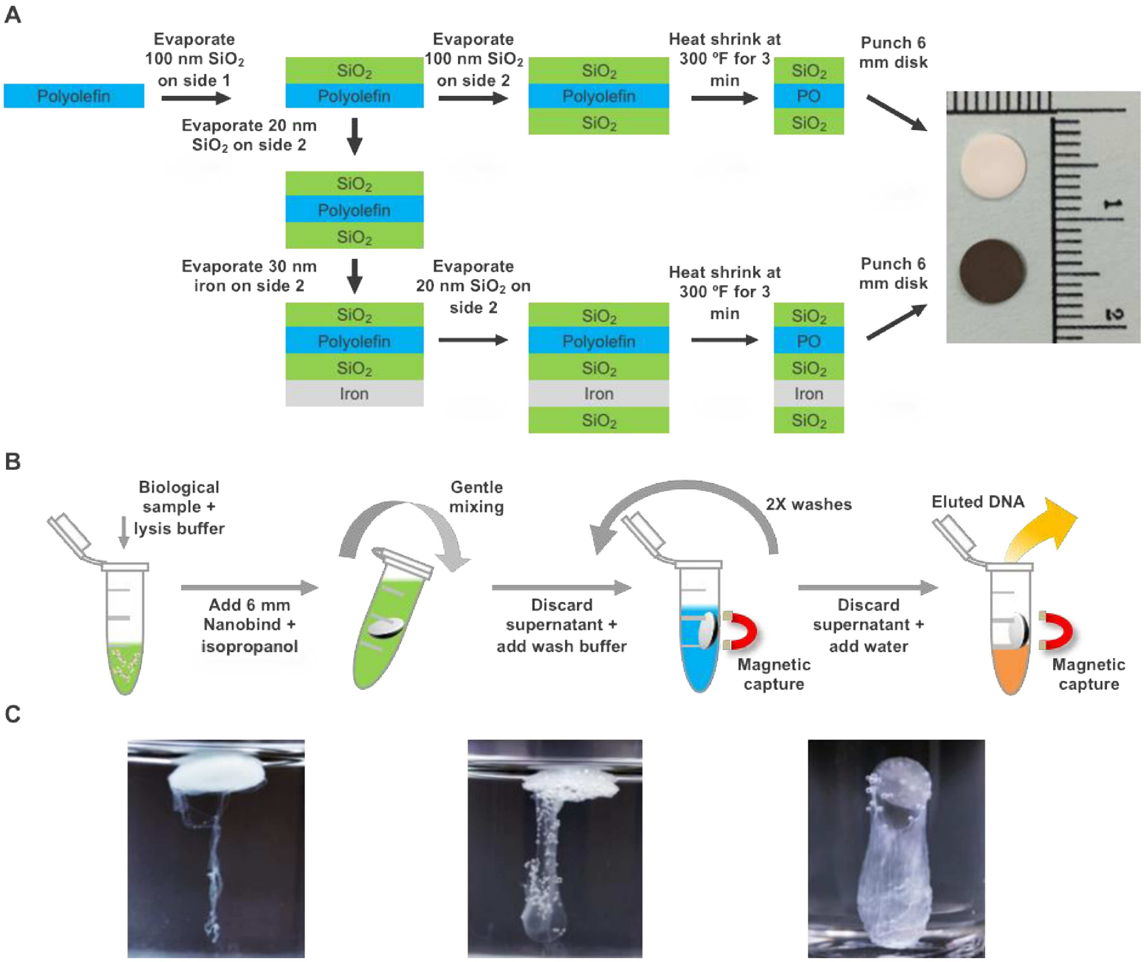

A
Polyolefin
Evaporate 100 nm SiO_2 on side 1
SiO_2
Polyolefin
Evaporate 100 nm SiO_2 on side 2
SiO_2
Polyolefin
SiO_2
Heat shrink at 300 °F for 3 min
SiO_2
PO
SiO_2
Punch 6 mm disk
1
2
Evaporate 20 nm SiO_2 on side 2
SiO_2
Polyolefin
SiO_2
Evaporate 30 nm iron on side 2
SiO_2
Polyolefin
SiO_2
Iron
Evaporate 20 nm SiO_2 on side 2
SiO_2
Polyolefin
SiO_2
Iron
SiO_2
Heat shrink at 300 °F for 3 min
SiO_2
PO
SiO_2
Iron
SiO_2
Punch 6 mm disk
B
Biological sample + lysis buffer
Add 6 mm Nanobind + isopropanol
Gentle mixing
Discard supernatant + add wash buffer
2X washes
Magnetic capture
Discard supernatant + add water
Eluted DNA
Magnetic capture
C

Figure 2.1. High molecular weight DNA extraction using the magnetic thermoplastic Nanobind

(A) The hierarchical silica lamella are fabricated by depositing a thin layers of silica onto a thermoplastic polyolefin (PO) film, heat shrinking in an oven, and punching out a 6 mm disk of material. Alternatively, an silica-iron-silica sandwich may be deposited on the backside to facilitate magnetic processing. The white disk is a non-magnetic Nanobind substrate while the dark gray disk is a magnetic Nanobind substrate. (B) DNA extraction is performed using a standard bind, wash, and elute protocol that can be performed with or without magnetic handling. (C) Images of genomic DNA captured by non-magnetic Nanobind and tethered onto silica surface after cell lysis and mixing with IPA. The high surface area and low shear forces enable large amounts of high molecular weight DNA to be captured via a tethered binding mode where macromolecular tentacles of DNA extend from the Nanobind substrate deep into solution. From left to right are images of Nanobind disks bound with genomic DNA from low (2x106), moderate (7x106), and high (25x106) cell inputs. Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

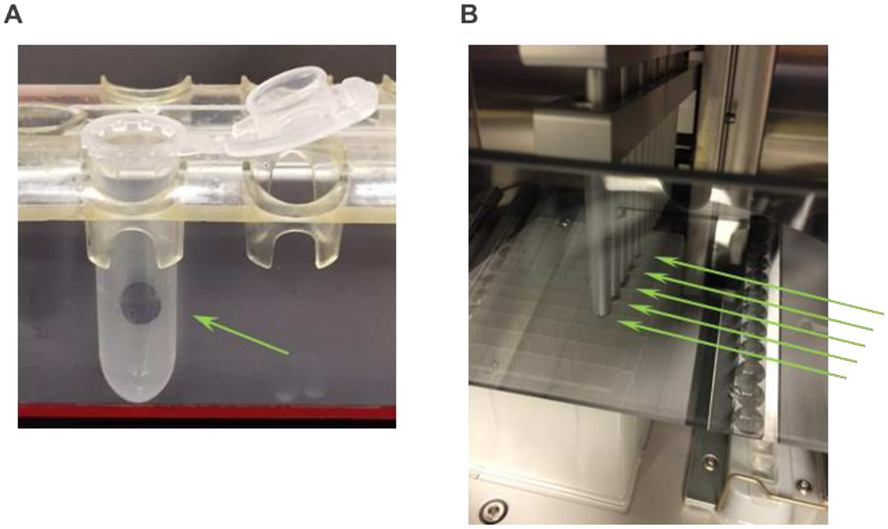

A
B

Figure 2.2. Images of magnetic Nanobind being used for manual and automated process

Images of magnetic Nanobind being used for (A) manual processing using a Life Technologies DynaMag-2 magnetic tube rack and (B) automated processing using a Thermo Scientific KingFisher DUO. The green arrows illustrate the black magnetic Nanobind substrates attracted to the tube wall and to the end of the magnetic rods, respectively. Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

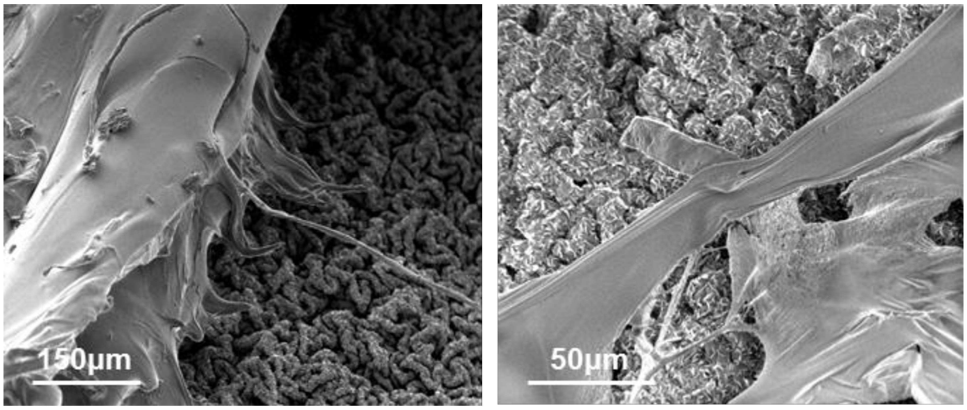

150µm
50µm

Figure 2.3. SEM images of genomic DNA bound to Nanobind

SEM images of genomic DNA bound to silica lamellae on a 100 nm oxide Nanobind substrate.

Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

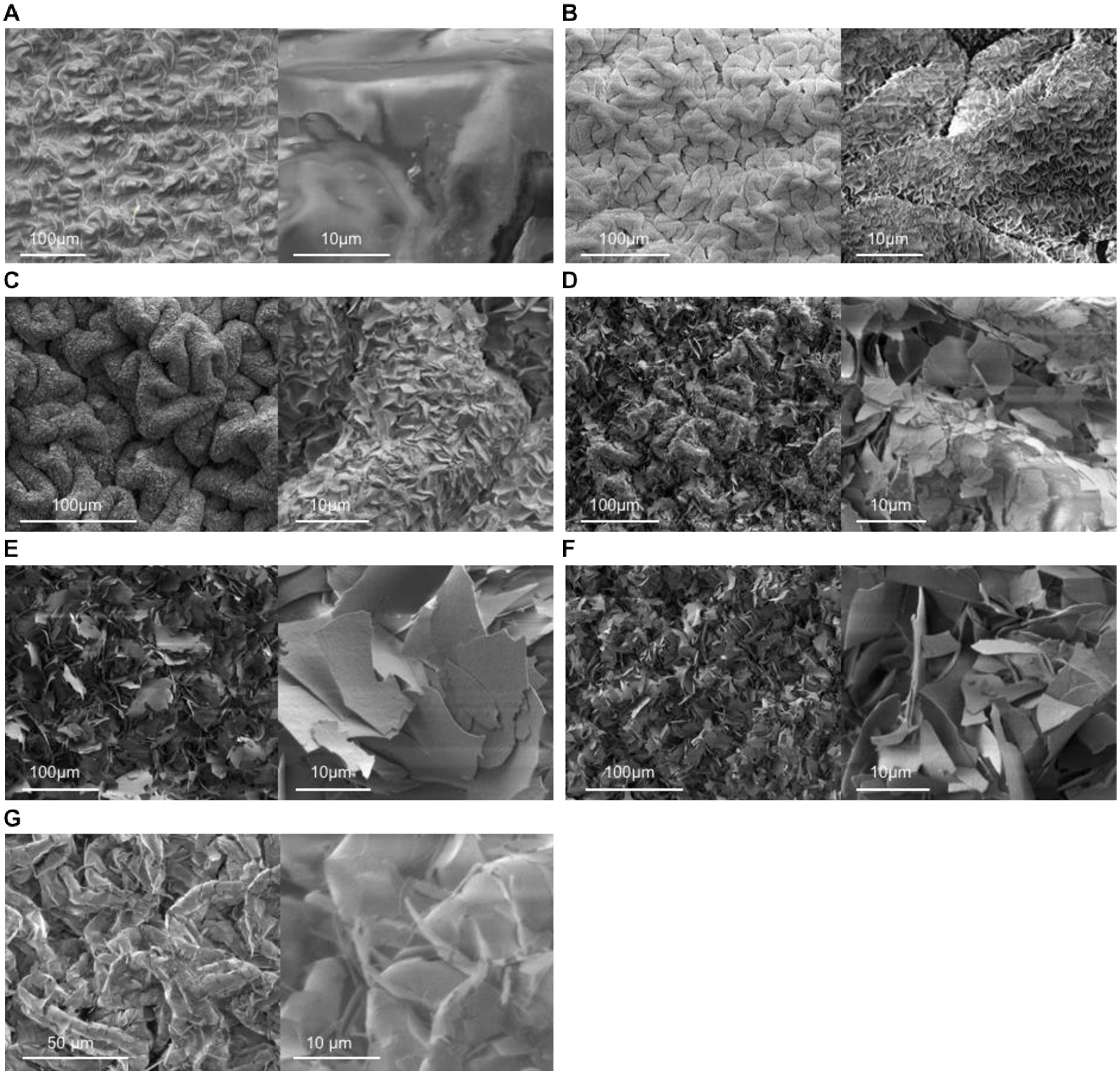

A
100µm
10µm
B
100µm
10µm
C
100µm
10µm
D
100µm
10µm
E
100µm
10µm
F
100µm
10µm
G
50 µm
10 µm

Figure 2.4. SEM images of Nanobind surface topography

The heat-shrinking process leads to a hierarchical structure of microscale polyolefin folds layered with nanoscale silica lamella that can be fine-tuned via oxide deposition thickness. SEM analysis was performed on non-magnetic Nanobind with (A) 2 nm, (B) 20 nm, (C) 50 nm, (D) 100 nm, (E) 150 nm and (F) 200 nm of deposited silicon dioxide. g) SEM images of magnetic Nanobind show a similar structure of microscale folds and nanoscale lamella. Each pair of images shows the same material under low **(left)** and high magnification **(right)**. Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

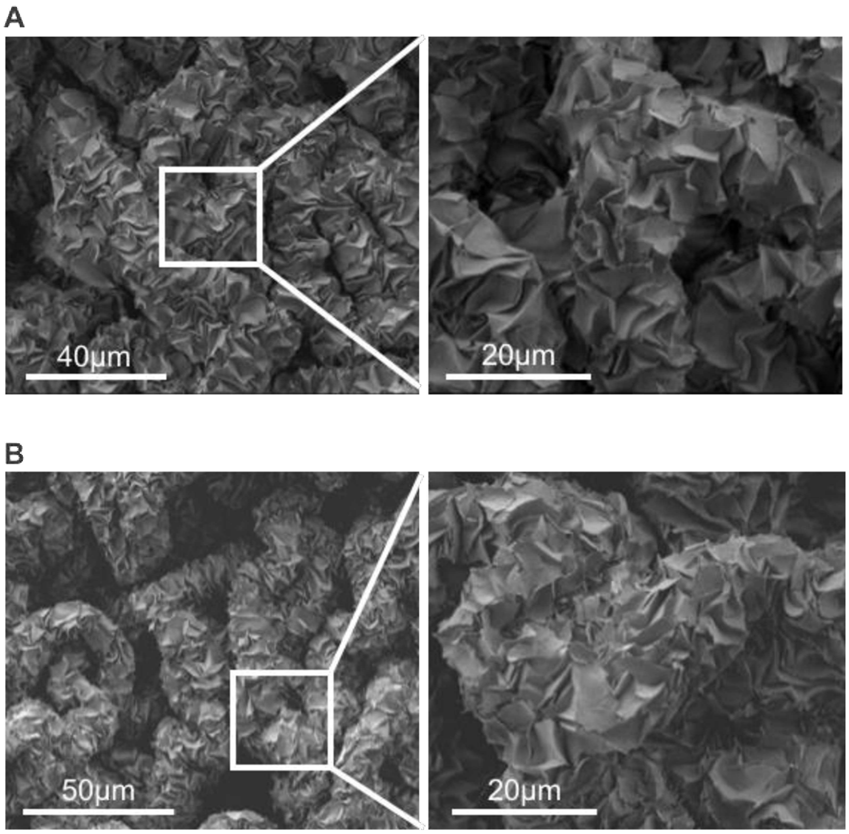

A
40μm
20μm
B
50μm
20μm

Figure 2.5. SEM images of Nanobind before and after DNA extraction

SEM images of a 100 nm oxide Nanobind substrates (A) before and (B) after DNA extraction. No change is seen in the surface topography demonstrating that the lamella are stably tethered. Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

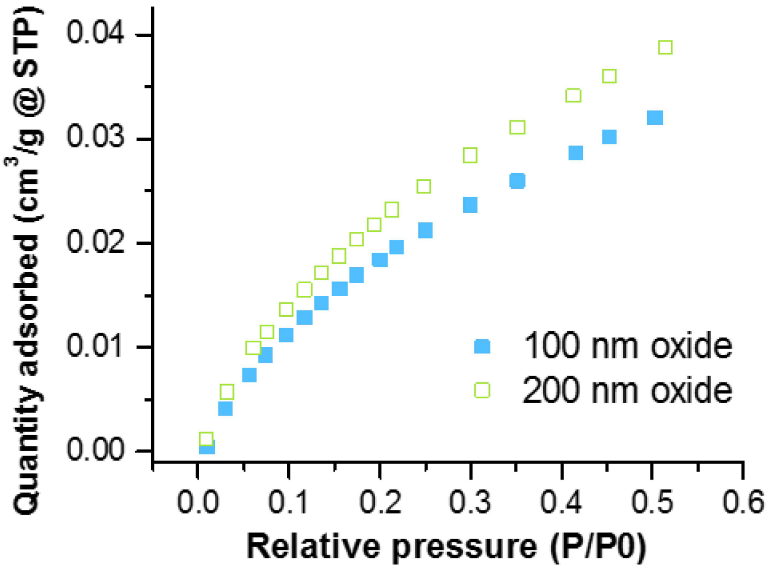

Quantity adsorbed (cm^3/g @ STP)
0.04
0.03
0.02
0.01
0.00
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Relative pressure (P/P0)
■ 100 nm oxide
□ 200 nm oxide

Figure 2.6. BET measurements for Nanobind

BET absorption isotherms for the 100 nm oxide and 200 nm oxide Nanobind substrates. Analysis resulted in BET surface areas of 9.33 cm^2 and 9.74 cm^2, respectively. Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

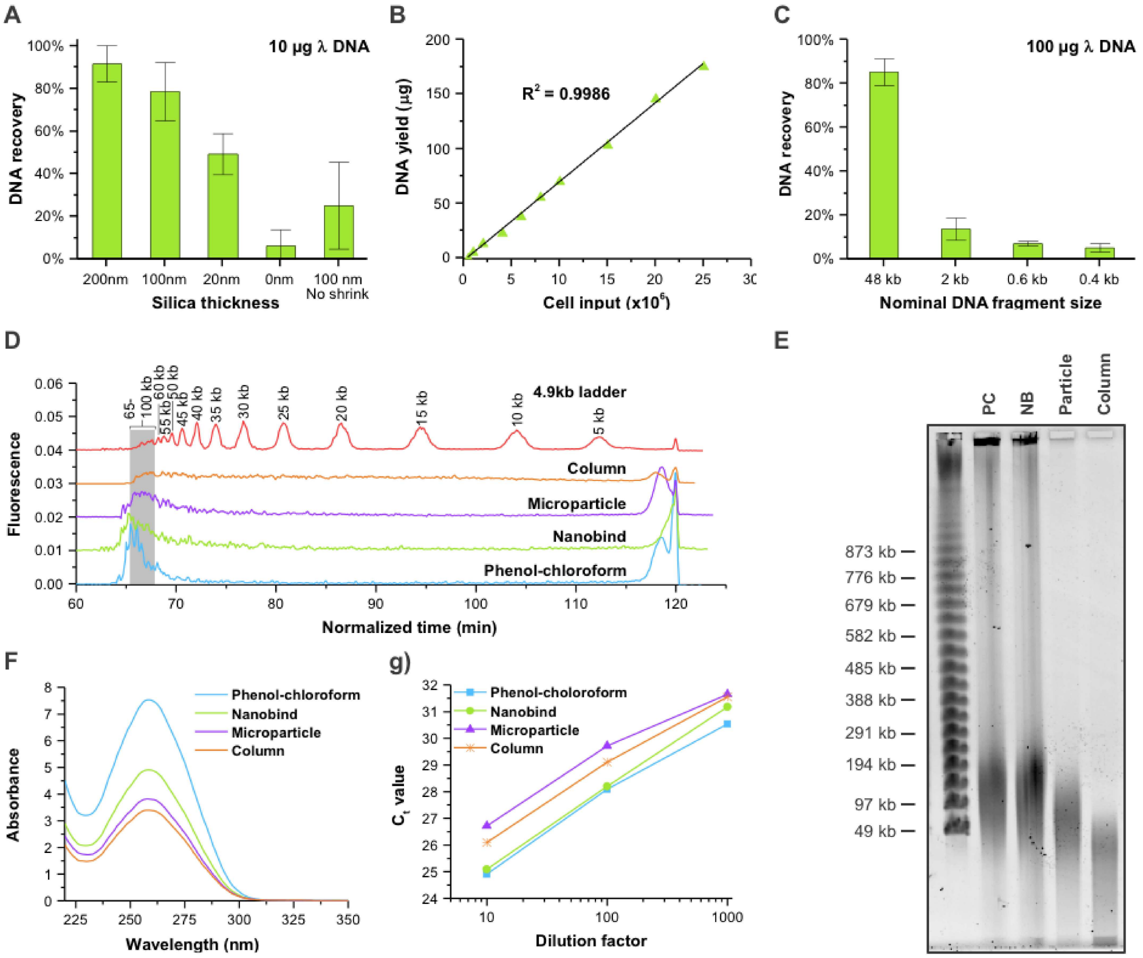
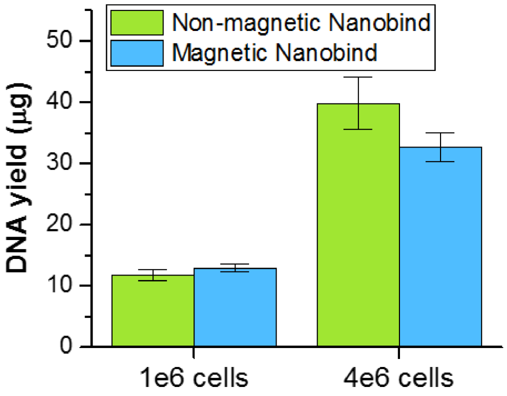

A
10 µg λ DNA
DNA recovery
100%
80%
60%
40%
20%
0%
200nm
100nm
20nm
0nm
100 nm
No shrink
Silica thickness
B
DNA yield (µg)
R² = 0.9986
200
150
100
50
0
0
5
10
15
20
25
30
Cell input (x10⁶)
C
100 µg λ DNA
DNA recovery
48 kb
2 kb
0.6 kb
0.4 kb
Nominal DNA fragment size
D
Fluorescence
0.06
0.05
0.04
0.03
0.02
0.01
0.00
65-100 kb
60 kb
55 kb
50 kb
45 kb
40 kb
35 kb
30 kb
25 kb
20 kb
15 kb
10 kb
5 kb
4.9kb ladder
Column
Microparticle
Nanobind
Phenol-chloroform
60
70
80
90
100
110
120
Normalized time (min)
E
PC
NB
Particle
Column
873 kb
776 kb
679 kb
582 kb
485 kb
388 kb
291 kb
194 kb
97 kb
49 kb
F
Absorbance
Phenol-chloroform
Nanobind
Microparticle
Column
225
250
275
300
325
350
Wavelength (nm)
g)
Ct value
Phenol-choloroform
Nanobind
Microparticle
Column
10
100
1000
Dilution factor

Figure 2.7. Nanobind DNA extraction characterization and comparison

(A) Extraction efficiency was measured as a function of Nanobind silicon dioxide thickness. The 200 nm and 100 nm silica Nanobind substrates contain a high density of silica lamella and achieve high extraction efficiencies when compared to the 0 nm oxide and planar oxide controls. (n=4) (B) DNA extraction yield was measured as a function of cell input. No saturation was seen even at 25×10^6 cells input, demonstrating that the binding capacity exceeds 188,000 $\mu g/m^2$. (C) Nanobind binding capacity was measured as a function of DNA length. Binding capacity was significantly enhanced for full length λ DNA versus highly fragmented λ DNA. (n=3) (D) Single molecule free solution hydrodynamic separation and (E) pulsed field gel electrophoresis were performed on genomic DNA extracted using phenol-chloroform (PC), Nanobind (NB), magnetic microparticles and spin column. Red dotted lines represent the median length of DNA extracted with each method. Phenol-chloroform and Nanobind DNA have the highest molecular weight with average length around 150 – 200 kb. Spin column DNA is highly fragmented (<50 kb) while magnetic particles show moderate shearing with DNA with average length around 75 kb. (F) UV absorbance shows that all 4 methods obtain high purity nucleic acids. g) qPCR was performed on serial dilutions of the extracted DNA. Nanobind and phenol-chloroform DNA consistently resulted in the lowest C_t values. (n=3). Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.



	1x10^6 cells				4x10^6 cells			
	Total nucleic acids (µg)	DNA (µg)	A260/ A280	A260/ A230	Total nucleic acids (µg)	DNA (µg)	A260/ A280	A260/ A230
Non-magnetic Nanobind	38±1	12±1	2.02±0.1	1.98±0.1	147±8	40±4	2.06±0.2	1.95±0.1
Magnetic Nanobind	43±3	13±1	1.95±0.1	2.00±0.1	144±17	33±2	2.14±0.1	2.10±0.1

Figure 2.8. Comparison between non-magnetic and magnetic Nanobind

DNA extraction was performed on either 1x10^6 MCF-7 cells or 4x10^6 MCF-7 cells using non-magnetic and magnetic Nanobind (n=3). UV absorbance was used to determine purity and total nucleic acid yield while Qubit assays were used to determine DNA yield. Across extensive testing, no noticeable difference has been seen in yield, binding capacity, purity, or molecular weight. Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

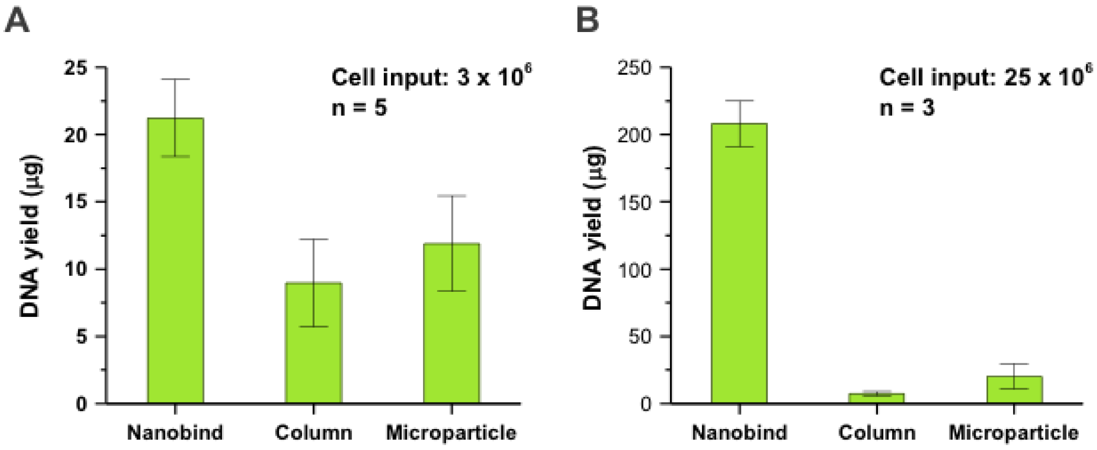

A
Cell input: 3 x 10^6
n = 5
DNA yield (µg)
0
5
10
15
20
25
Nanobind
Column
Microparticle
B
Cell input: 25 x 10^6
n = 3
DNA yield (µg)
0
50
100
150
200
250
Nanobind
Column
Microparticle

Figure 2.9. Yield comparison of Nanobind, spin column, and magnetic microparticles

(A) With $3x10^6$ cells input, Nanobind showed 2-fold higher yield as spin column and microparticles were already saturated. (B) At $25x10^6$ cells input, Nanobind still had not reached saturation and showed 10-fold higher yield than microparticles and nearly 30-fold higher yield than microparticles. Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

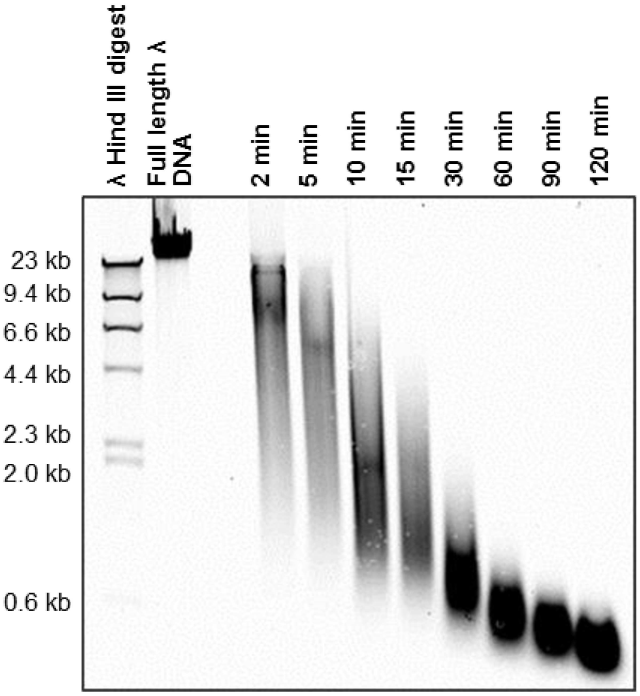

λ Hind III digest
Full length λ DNA
2 min
5 min
10 min
15 min
30 min
60 min
90 min
120 min
23 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
0.6 kb

Figure 2.10. Gel image of heat sheared λ DNA used for DNA length analysis

Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

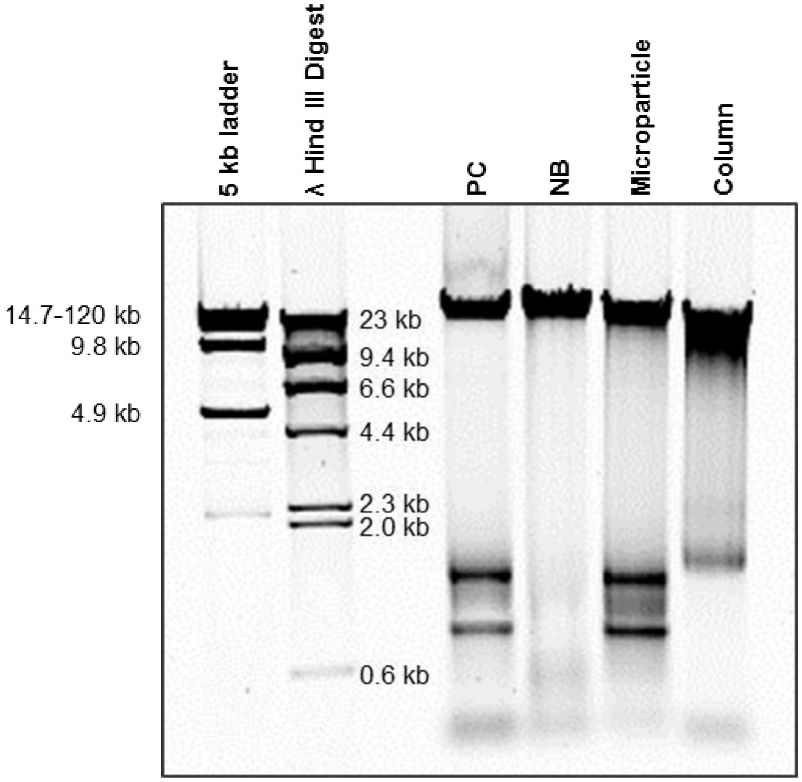

5 kb ladder
λ Hind III Digest
PC
NB
Microparticle
Column
14.7-120 kb
9.8 kb
4.9 kb
23 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
0.6 kb

Figure 2.11. Gel image of DNA extracted using four different methods

Agarose gel electrophoresis image of DNA extracted using phenol-chloroform (PC), Nanobind (NB), magnetic microparticles, and spin column. Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

Extraction method	260/280	260/230
Phenol-chloroform	1.991±0.007	2.150±0.005
Nanobind	2.000±0.008	2.197±0.032
Microparticle	2.041±0.010	2.095±0.075
Spin column	2.041±0.024	2.251±0.036

Figure 2.12. UV aborbance curves of nucleic acid extracted using four methods

UV absorbance curves of nucleic acids extracted from 3 million cells by phenol-chloroform, Nanobind, magnetic microparticles, and spin column. All 4 methods obtain high purity nucleic acids. The extracted nucleic acids contain a mixture of DNA and RNA. Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

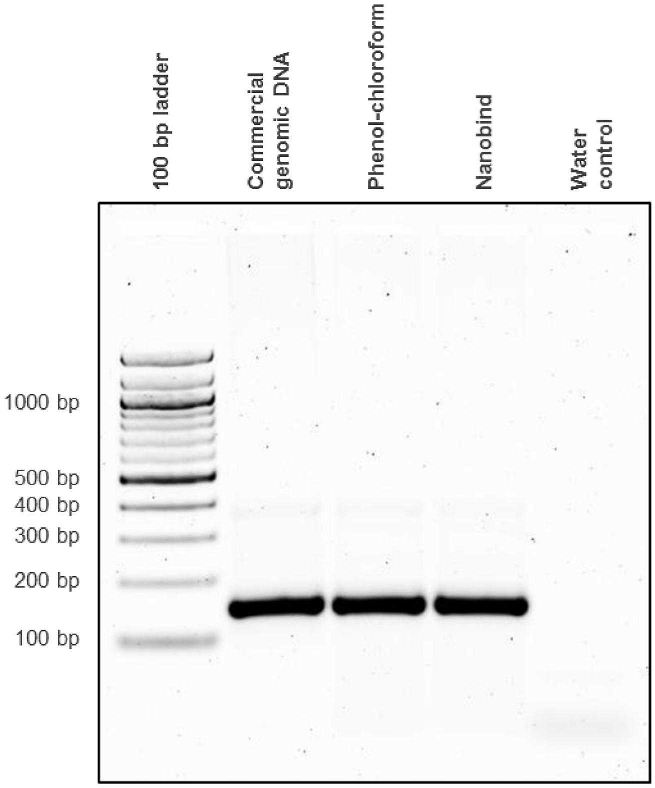

100 bp ladder
Commercial genomic DNA
Phenol-chloroform
Nanobind
Water control
1000 bp
500 bp
400 bp
300 bp
200 bp
100 bp

Figure 2.13. Gel image of PCR amplification products

Agarose gel electrophoresis image of the 148 bp PCR product amplified from commercial genomic DNA, phenol-chloroform extracted DNA, Nanobind extracted DNA, and a water control. Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

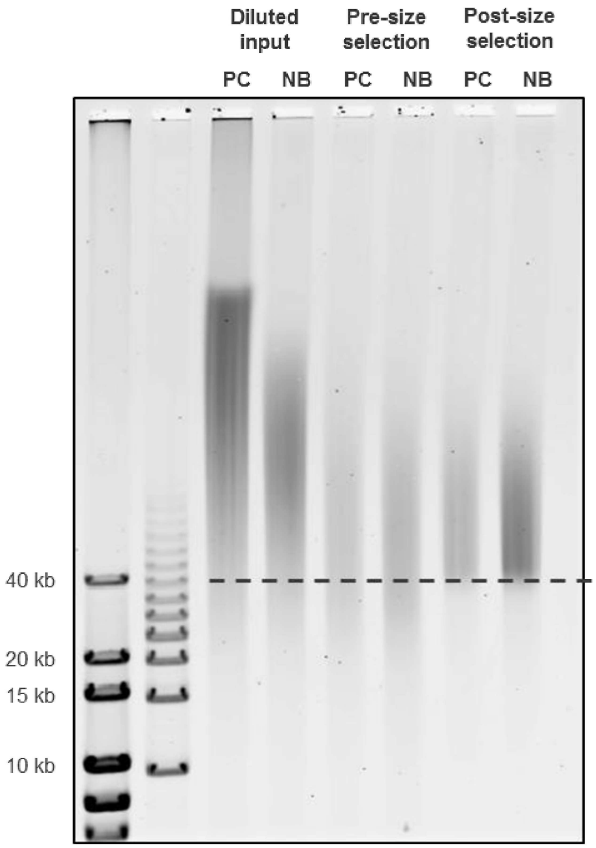
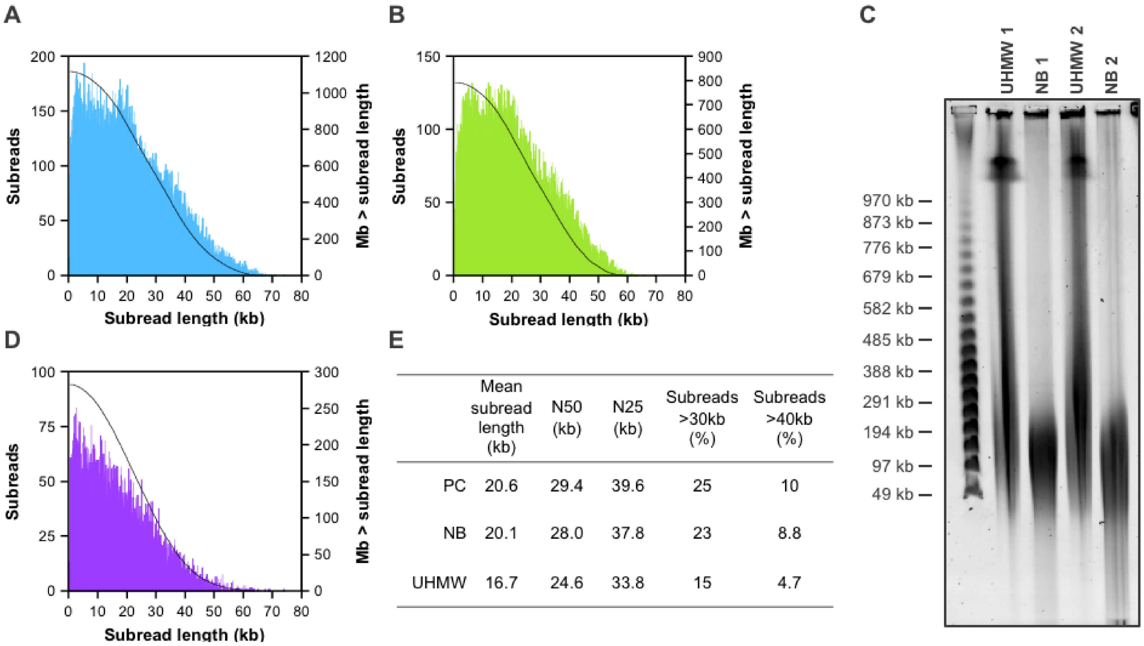

Diluted input
Pre-size selection
Post-size selection
PC
NB
PC
NB
PC
NB
40 kb
20 kb
15 kb
10 kb

Figure 2.14. Phenol-chloroform and Nanobind DNA SMRTbell libraries size selection

Phenol-chloroform (PC) and Nanobind (NB) DNA SMRTbell libraries size selected using a 30 kb cutoff. Dashed line indicates the 40 kb mark. Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.



	Mean subread length (kb)	N50 (kb)	N25 (kb)	Subreads >30kb (%)	Subreads >40kb (%)
PC	20.6	29.4	39.6	25	10
NB	20.1	28.0	37.8	23	8.8
UHMW	16.7	24.6	33.8	15	4.7

Figure 2.15. SMRT sequencing data for phenol-chloroform, Nanobind and UHMW Nanbind DNA

>30 kb insert SMRTbell libraries prepared using a) PC and b) NB extracted DNA resulted in exceptional read lengths with N50 values of 29.2 kb and 27.7 kb, respectively. c) An UHMW extraction process was developed that increased extracted DNA size from 300 kb up to Mb range. d) The UHMW DNA was then used to generate >30 kb insert SMRTbell libraries that resulted in robust long reads with N50 values of 24 kb and longest reads of 70 kb. e) SMRT sequencing read metrics for all libraries. All data is based on the max subread length obtained per zero-mode waveguide. Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

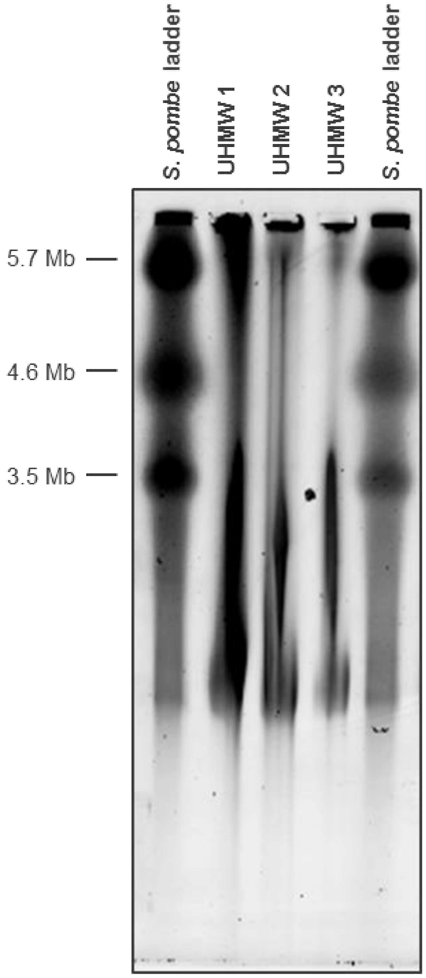

S. pombe ladder
UHMW 1
UHMW 2
UHMW 3
S. pombe ladder
5.7 Mb
4.6 Mb
3.5 Mb

Figure 2.16. Gel image of UHMW Nanobind DNA under PFGE condition

Ultra high molecular weight (UHMW) Nanobind DNA from 3 separate extractions was run under alternate PFGE conditions to more accurately determine DNA size. When compared to the *S. pombe* ladder, UHMW DNA can be seen surpassing 5.7 Mb in size. Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

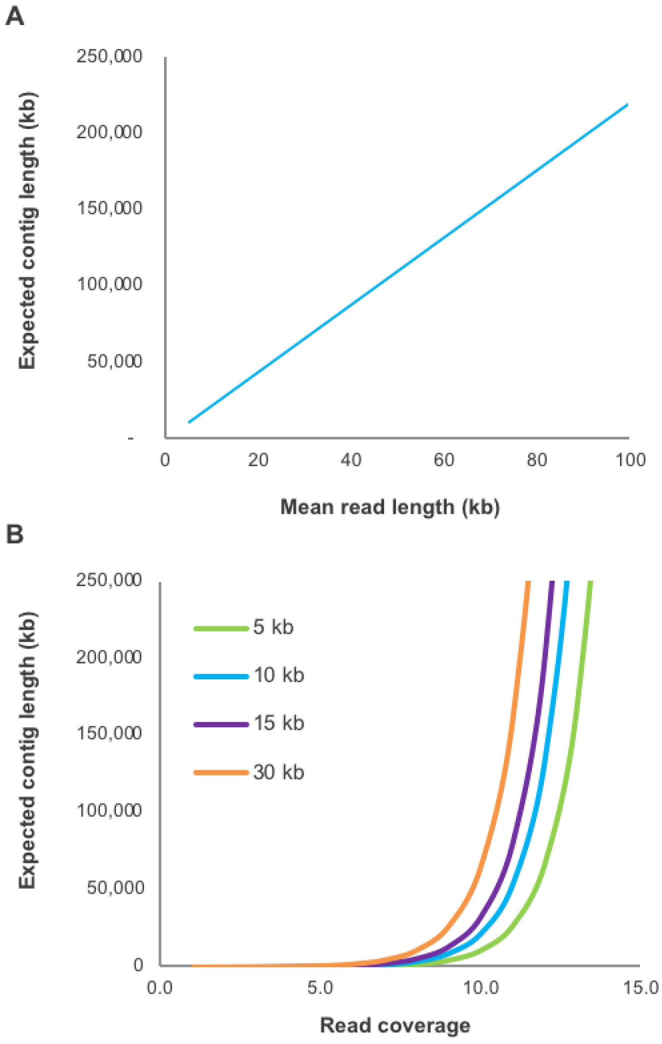

A
Expected contig length (kb)
250,000
200,000
150,000
100,000
50,000
-
0
20
40
60
80
100
Mean read length (kb)
B
Expected contig length (kb)
250,000
200,000
150,000
100,000
50,000
0
5 kb
10 kb
15 kb
30 kb
0.0
5.0
10.0
15.0
Read coverage

Figure 2.17. Expected contig length vs read length or coverage by Lander-Waterman model a) Expected average contig length as a function of mean read length for 10X coverage. b) Expected average contig length as a function of read coverage across a range of mean read lengths for the human genome. Calculations are based on the Lander-Waterman model. Adapted with permission from [91]. Copyright 1999-2019 John Wiley & Sons, Inc.

s

Chapter 3

3. Ratiometric Fluorescence Coding Assay for Multiplexed Nucleic Acid Detection

3.1 Background

Nucleic acid amplification testing (NAAT) has emerged as a popular technique for the diagnosis of many diseases including cancer [43, 44], genetic disorders [92], mitochondrial disorders [93], and infectious diseases [42, 94, 95]. Indeed, NAAT methods such as real-time polymerase chain reaction (PCR) [45, 46], ligase chain reaction (LCR) [96, 97], and nucleic acid sequence-based amplification (NASBA) [98, 99] have been widely employed for nucleic acid detection because of their high sensitivity, specificity, and rapid turn-around-time. In addition, a number of different approaches can be coupled with NAAT to achieve multiplexed detection of multiple targets in a single assay, which is a particularly useful capability for clinical diagnosis of diseases [94, 100]. Among these approaches, which range from differentiating the melting temperatures of NAAT amplicons [51, 52] to assigning targets with probes that have distinct lengths [49, 50, 101, 102], the most commonly employed strategy has been using a target-specific, fluorescently-labeled oligonucleotide probe for each target nucleic acid sequence. While this "one-color-one-target" approach is straightforward and effectively leverages the mature fluorescence detection infrastructure associated with NAAT methods, it unfortunately has a limited capacity of multiplexing because spectral overlaps between fluorophores restrict the number of fluorophores that can be used within a single assay. As such, there remains a need for fluorescence-based approaches that extend beyond the one-color-one-target restriction and expand the multiplexing capacity of NAAT.

A natural and promising extension to the one-color-one-target approach is using a unique combination of multiple fluorophores to encode each target sequence. A handful of strategies for demonstrating this approach have been reported. For example, 15-plex detection using only 4 fluorophores has been achieved through a strategy called Multicolor Combinational Probe Coding (MCPC) [103, 104]. To achieve 15-plex detection via MCPC, however, 32 different fluorescently-labeled probes were synthesized, which escalated the cost of this strategy. Moreover, the probe design principle in MCPC capped its multiplexing capacity to 15 when using 4 fluorophores; and as spectral overlaps between fluorophores make it difficult to incorporate additional fluorophores, MCPC had likely reached its maximum multiplexing capacity. Another strategy, which was coined binary-scheme mathematical theory, employed each fluorescence color as a “digit” in the binary code and assigned each target sequence with a unique binary code of at least two colors/digits [105]. Unfortunately, even when using 4 fluorophores, this strategy still only achieved 4-plex detection. Therefore, despite the potential, effective strategies for leveraging combinations of multiple fluorophores to achieve multiplexed NAAT remain underdeveloped to date.

Herein, we describe Ratiometric Fluorescence Coding, our new strategy for expanding the multiplexed detection capacity of NAATs by encoding each nucleic acid target with a specific fluorescence ratio. Specifically, in Ratiometric Fluorescence Coding, we transform each target into a specific DNA template that can hybridize with multiple fluorescently-labeled probes with distinct colors at a pre-designed, unique ratio. In doing so, we can specifically detect each target while obviating the restriction of spectral overlaps between fluorophores. Moreover, since only the ratios of fluorescent probe hybridization sites in DNA templates are designed to vary but the sequences of the hybridization sites remain the same, our Ratiometric Fluorescence Coding can utilize the same set of fluorescently-labeled probes to detect multiple targets. For demonstrating the concept of Ratiometric Fluorescence Coding, we utilized padlock probe chemistry [106] for probe design and coupled with either rolling circle amplification (RCA) or hyperbranched rolling circle amplification (HRCA) [107-109] for DNA template generation, which allows hybridization with fluorescent-labeled probes. Using our RCA and HRCA assays and only two fluorescently-labeled probes, we demonstrate the detection of the gene sequences from up to 6 infectious diseases, including 4 sexually transmitted diseases (STDs) and 2 urinary tract infections (UTIs), with distinct and highly resolved fluorescence ratios, illustrating that Ratiometric Fluorescence Coding offers ready means for expanding the multiplexed detection capacity of NAATs. In addition, we have coupled our Ratiometric Fluorescence Coding assay with a droplet-based microfluidics system to extend its utility to simultaneously detecting multiple targets in a single reaction, termed Digital Ratiometric Fluorescence Coding. For demonstration, we have detected DNA targets from six sexually transmitted diseases and demonstrated its potential for further expanding the multiplexing capability of our method.

3.2 Experimental Details

3.2.1 Materials and Reagents

All DNA oligonucleotides including synthetic targets, padlock probes, and primers were purchased from Integrated DNA Technologies, Inc. (Coralville, IA) and both fluorescently-labeled peptide nucleic acid (PNA) probes were purchased from PNA Bio (Newbury Park, CA). Reagents for ligation, RCA, and HRCA reactions, including 9°N™ DNA Ligase, 10× 9°N™ DNA Ligase Reaction Buffer, Phage T4 Gene-32 Protein, Phage ø29 DNA Polymerase, Exonuclease I (*E. coli*), Exonuclease III (*E. coli*), 10× Isothermal Reaction Buffer, Bst DNA polymerase, and 10× Bst Polymerase Reaction Buffer were purchased from New England BioLabs, Inc. (Ipswich, MA). Deoxyribonucleotide triphosphate (dNTP) was purchased from Invitrogen Corp. (Carlsbad, CA).

Formamide, sodium chloride (NaCl), EDTA (pH 8.0), Bovine Serum Albumin (BSA), Betaine and H_2O (PCR grade) were purchased from Sigma-Aldrich Corp. (St. Louis, MO).

3.2.2 Nucleic Acid Oligonucleotide Sequences

Padlock probes targeting gene sequences in *Proteus mirabilis* (PM, 238:275 from AF240693) *Human immunodeficiency virus 1* (HIV-1, 258:305 from KC966998), *Neisseria gonorrhoeae* (NG, 362:411 from X52364), *Chlamydia trachomatis* (CT, 50:98 from JX648604), *Treponema pallidum* (TP, 801:846 from KC966998), and *Escherichia coli* (EC, 189:230 from AY447194) were designed in-house. All padlock probe and synthetic target sequences are given in **Table 3.1**. The sequence of the RCA primer is 5'-CTAAAGCTGAGACATGACGAGTC and the sequence of the additional primer for performing HRCA is 5'- TCAGAACTCACCTGTTAG. The sequences of the red and the green PNA probes are [5’AlexaF647] TCAGAACTCACCTGTTAG and [5’AlexaF488] CCCTAACCCTAACCCTAA, respectively.

For droplet-based Digital Ratiometric Fluorescence Coding assay, another group of six padlock probes Padlock probes targeting gene sequences in *Microplasma genitalium* (MG, 1420:1473 from M31431), *Herpes Simplex Virus* (HSV, 1426:1473 from K01760), *Neisseria gonorrhoeae* (NG, 362:411 from X52364), *Human immunodeficiency virus* (HIV, 258:305 from KC966998), *Treponema pallidum* (TP, 801:846 from KC966998) and *Chlamydia trachomatis* (CT, 50:99 from JX648604) were designed in-house. All padlock probe and synthetic target sequences are given in **Table 3.2**. The sequence of the HRCA primers are 5'-TTCGTTCGTTTATCGGGTATTT TAGT and 5'- CGATACCGAACCTAAAACTAATC. The sequences of the red and the green MBs are /5Alex555/cCGaCtCaAcCGG/3BHQ_2/ and /5Alexa488/tCCgAgCgTaGGA/3BHQ_1/, respectively (Locked Nucleic Acid (LNA) bases are denoted in lower case). All oligonucleotides were purchased from Integrated DNA Technologies, Inc. (Coralville, IA).

3.2.3 Padlock Probe Hybridization and Ligation

Our Ratiometric Fluorescence Coding assays began with padlock probe hybridization and ligation, which were achieved in a single-tube reaction. The 25-µL reaction mixture contained 45 nM of each linear padlock probe, synthetic DNA targets (at different concentrations), 0.4 Units/µL 9°N™ DNA Ligase, and 1× 9N DNA Ligase Reaction Buffer (10 mM Tris-HCl, 600 µM ATP, 2.5 mM $MgCl_2$, 2.5 mM Dithiothreitol, 0.1% Triton X-100, pH 7.5 at 25°C). Padlock probe hybridization and ligation was performed at 60 °C for 1 hour. For subsequent HRCA reaction, only 100 pM of each padlock probe were ligated in the presence of synthetic DNA targets or pathogen genomic DNA to reduce non-specific amplification and improve signal-to-noise-ratio.

The padlock probe hybridization and ligation reactions for Digital Ratiometric Fluorescence Coding assay was performed in a 8-µL reaction mixture containing 50 pM of each linear padlock probe, synthetic DNA targets at different concentrations, 1.5 Units/µL 9°N™ DNA Ligase, and 1× 9N DNA Ligase Reaction Buffer. Padlock probe hybridization and ligation was performed at 55 °C for 1 hour. The 8-µL ligation product was then treated with 10 units of Exonuclease I and 50 units of Exonuclease III at 37 °C for 2 hour to remove extra linear padlock probes that were not circularized, followed by a brief incubation at 85 °C for 10 minutes to inactivate the enzymes.

3.2.4 RCA and HRCA Reactions

RCA or HRCA were performed immediately after ligation of padlock probes to generate DNA templates that allow hybridization with fluorescently-labeled probes. For RCA, 2 µL of the ligation product was mixed with 800 nM rolling-circle primer, 400 µM dNTPs, 10 U/µL Phage ø29 DNA polymerase, and 1× Isothermal Reaction Buffer (50 mM Tris•HCl (pH 7.5), 10 mM $MgCl_2$, 200 µg/mL acetylated BSA) to a final volume of 25 µL. Of note, due to its high strand-displacing activity, Phage ø29 DNA polymerase was added last to the reaction mixture, which was then immediately incubated at 31 °C for 1 hour to perform RCA, followed by 70 °C for 10 minutes to inactivate the polymerase and stop the reaction.

For HRCA, 25 μL of the ligation product was first treated with 10 units of Exonuclease I and 50 units of Exonuclease III at 37 °C for 1 hour to remove extra linear padlock probes that were not circularized, followed by a brief incubation at 85 °C for 10 minutes to inactivate the enzymes. Then 4 μL of this enzyme-treated ligation product was mixed with 400 nM of each of the two rolling-circle primers, 400 μM dNTPs, 40 ng/ml Phage T4 Gene-32 Protein, 10 U/μL Bst DNA polymerase, and 1× Bst Polymerase Reaction Buffer (20 mM Tris•HCl (pH 8.8), 10 mM KCl, 2.7 mM $MgSO_4$, 5% v/v DMSO, 0.1% Triton × 100) to a final volume of 25 μL. The reaction mixture was incubated at 60 °C for 1 hour to perform HRCA, followed by 85 °C for 10 minutes to inactivate the polymerase and stop the reaction.

3.2.5 Fluorescent PNA Hybridization and Detection

10 μL RCA or HRCA product was mixed with 500 nM AlexaF488-labeled PNA probe and 500 nM AlexaF647-labeled PNA probe in a PNA hybridization buffer (40% Formamide, 10 mM NaCl, and 50 mM Tris•HCl (pH 8)), heated at 85 °C for 5 minutes, and then incubated at room temperature in the dark for 2 hours to allow hybridization. Non-hybridized, excess fluorescent PNA probes were removed by using S400-HR microspin columns (GE Healthcare). Each purified sample was pipetted into a well of a 384-well plate (Corning Inc.) and read out on a Typhoon image scanner (GE Healthcare) in both the red channel (peak excitation at 633 nm with 670 nm band-pass emission filter) and the green channel (peak excitation at 488 nm with 526 nm short-pass emission filter) at a photomultiplier tube (PMT) voltage of either 600 V or 650 V. Of note, for each experiment, the PMT voltage was kept the same between the red channel and the green channel (i.e., both at 600 V or both at 650 V). Between experiments, however, the PMT voltage was sometimes adjusted between either at 600 V or at 650 V.

3.2.6 Microfluidic Chip Fabrication

Our microfluidic device was featured with a 10-mm flow focusing junction. The mold for our device was first prepared via soft lithography. For chip fabrication, the mold was initially silanized with chlorotrimethylsilane (Sigma-Aldrich, St. Louis, MO, USA) for 15 min before pouring PDMS (SYLGARD 184 Silicone Elastomer Kit, Dow Corning, Midland, MI, USA) onto the mold. After curing for 1 hour at 80°C, the 5mm PDMS slab was peeled off, on which the holes were punched using a sharpened needle (20 gauge; McMaster-Carr, Elmhurst, IL, USA) as inlets and outlets. Then the PDMS slab was cleaned by a thorough hole cleaning using IPA to eliminate the PDMS residues. The cleaned PDMS slab was bonded onto a glass slide using oxygen plasma treatment before hydrophobic treatment with Aquapel.

3.2.7 Droplet Generation and Amplification

For our Digital Ratiometric Fluorescence Coding, 4 μL of enzyme-treated ligation product was mixed with 400 nM of each of the two HRCA primers, 200 nM of each of the two MBs, 400 μM dNTPs, 0.25 mg/mL BSA, 800 nM Betaine, 40 ng/ml Phage T4 Gene-32 Protein, 0.32 U/μL Bst DNA polymerase, and 1× Bst Polymerase Reaction Buffer (NEB) to a final volume of 50 μL. The 50-μL HRCA reaction mixture was kept in ice before loaded for droplet generation.

We used microfluidic device to encapsulate the circularized padlock probes into picoliter-volume droplets for digital HRCA reaction and ratiometric fluorescence coding. Basically, the circularized probe encapsulation in droplets follows the Poisson distribution. In order to separate single circularized probe and achieve multiplexing detection, the circularized padlock probes were appropriately diluted to make sure that most of droplets encapsulate either zero or single copy probe for HRCA amplification so that the produced color ratio was not messed up by each other.

Specifically, the microfluidic device employed a 10-μm flow-focusing junction to generate 3-pL droplets with one inlet for droplet generation oil (008-FluoroSurfactant dissolved HFE 7500, RAN Biotechnologies, Inc., Beverly, MA) and two inlets for amplification mixture. To generate droplets, the 50-µL HRCA reaction mixture was loaded into two Tygon tubings which connected to a Hamilton 1000 glass syringe with FC40 oil as a displacement fluid and then injected into the microfluidic device using a syringe pump (Harvard Apparatus, Holliston, MA) at a 50 µL/h flow rate. The droplet generation oil was driven at 300 µL/h by another separate syringe pump to shear the HRCA mixture into droplets at the 10-µm flow-focusing junction. Millions of droplets can be generated within 15 mins and they were then collected in a tube with extra oil removed to pack them. The droplets containing HRCA reaction mixture were incubated at 60 °C for 2 hours for digital HRCA amplification, followed by 95 °C for 5 minutes to inactivate the polymerase and denature double-strand oligonucleotides for MB hybridization. The green and red MBs then bind with the denatured amplified targets at 37 ° for 20 min as the last step before detection.

3.2.8 Droplet Fluorescence Measurement

To interrogate the fluorescence of single picoliter-volume droplets after MB binding with the amplified targets, a custom-built confocal microscope detection system with an adjustable stage was used for fluorescence detection. First, the amplified droplets were loaded into Tygon tubings with caution to avoid droplet merging and reinjected into the microfluidic device with a 15-µL/h flow rate. To achieve single droplet detection, the detection module of our device used a 10-µm constriction channel to get single droplet flow. Then two laser sources, 488 nm (2 mW) and 545 nm (6 mW) respectively, were focused in the middle of the 10-µm-constriction channel to excite the fluorophores. The red and green fluorescence intensities of each droplet flowing through the channel were measured with APDs with a 0.1 ms sampling time and recorded by a custom-written LabView program for subsequent analysis.

3.2.9 Data Analysis

For our Ratiometric Fluorescence Coding assay performed in tubes, red and green fluorescence intensities of each sample measured by the Typhoon image scanner were quantified by ImageQuantTL software (GE Healthcare). The R/G ratio was calculated via dividing the red fluorescence intensity by the green fluorescence intensity.

For our Digital Ratiometric Fluorescence Coding assay, the droplet fluorescence trace was loaded into MATLAB for data processing. First, droplets with appropriate size were selected to rule out any droplet merging or shearing during operation. Then a double gaussian fitting was used to fit the fluorescence intensity of the selected droplets to get the mean and standard deviation for empty droplets ('negative' droplets) which were used to set the threshold for selecting droplets with circularized padlock probe ('positive' droplets). The threshold for red channel was set to be equal to negative mean plus 9 folds of negative standard deviation to get rid of false positive signal. The threshold for green channel was set to be 225 since we noticed that there was some low non-specific amplification signal in green channel and they were typically smaller than 225. Droplets with fluorescence signal above the threshold of either green or red channel were considered to be positive. In order to calculate the color ratios for these positive droplets, we subtracted the negative mean from both channels to get the true signals and used the ratio between the true signals to characterize and identify different targets.

3.3 Design of Padlock Probes for Ratiometric Fluorescence Coding

Our Ratiometric Fluorescence Coding encodes each target of interest with a specific fluorescence color ratio. This is achieved by transforming each target to a specific DNA template sequence with a pre-designed and unique ratio of binding sites for multiple fluorescently-labeled probes with distinct colors. As such, upon hybridization with fluorescent probes, the template-probe complex displays the specific ratiometric fluorescence code whose intensity of each fluorescence color can be measured to decode and identify the target (**Figure 3.1A**). Importantly, in Ratiometric

Fluorescence Coding, the number of maximum target-specific fluorescence ratios (i.e., the multiplexing capacity) for a given number of fluorescence colors is governed by a polynomial relationship with the number of fluorescent probe binding sites (S) that can be designed on a DNA template. In the case of 2 fluorescence colors, the maximum achievable fluorescence ratios are $\frac{(S^2+4S-1)}{4}$ for an odd number of S and $\frac{(S^2+3S+3)}{4}$ for an even number of S, respectively. For example, with only a red fluorophore (e.g., AlexaF647) and a green fluorophore (e.g., AlexaF488), we can achieve up to 7 fluorescence ratios if we design 4 fluorescent probe binding sites (S = 4) on each DNA template, and up to 11 ratios if we design 5 binding sites (S = 5) on each DNA template. Moreover, in Ratiometric Fluorescence Coding, only the numbers of fluorescence probe binding sites in the DNA templates are designed to vary but the sequences of the fluorescence probe binding sites remain the same. As such, the same fluorescently-labeled probes can be used for detecting multiple targets. These features allow our Ratiometric Fluorescence Coding to expand the multiplexing capacity when compared with the traditional one-color-one-target approach and multi-fluorophore strategies without requiring a large number of fluorescently-labeled probes.

Padlock probe chemistry provides an effective means for achieving Ratiometric Fluorescence Coding due to its flexibility for probe design. In this work, we design each padlock probe to contain a target-specific hybridization region, a uniquely pre-designed ratio of binding sites for the red and the green fluorescent probes, and general primer sites (**Figure 3.1B**). In the presence of its specific DNA target, the initially linear padlock probe is first hybridized to the target and then ligated to become a circular probe, from which DNA templates can be generated via various amplification techniques to amplify the number of fluorescent probe binding sites at the pre-designed fluorescent probe ratio encoded in each padlock probe [107, 108, 110]. Following this design principle, we designed six circularizable padlock probes to detect 6 common infectious diseases (4 STDs and 2 UTIs) in this work. The target-specific regions of these 6 padlock probes target conserved sequences from *Proteus mirabilis* (PM), *Human immunodeficiency virus 1* (HIV-

1), *Neisseria gonorrhoeae* (NR), *Chlamydia trachomatis* (CT), *Treponema pallidum* (TP) and *Escherichia coli* (EC). Moreover, we incorporated the binding sites for the red and the green fluorescent probes at specific red to green (R/G) ratios into these padlock probes: 0:1 for PM, 1:2 for HIV-1, 1:1 for NR, 2:1 for CT, 3:1 for TP and 1:0 for EC, respectively (**Figure 3.1C**).

We have also coupled padlock probe chemistry with molecular amplification techniques, such as RCA and HRCA, to amplify the fluorescence signals in ratiometric fluorescence coding. More specifically, we begin by challenging the infectious disease-related DNA targets to the mixture of our six target-specific, linear padlock probes. In the presence of a DNA target, the corresponding linear padlock probe is hybridized and ligated to form a closed padlock probe. Then RCA reaction amplifies along the circular padlock probe to generate DNA templates with hundreds of tandem repeats that serve as hybridization sites for fluorescent probes. In this work, we employed AlexaF488- and AlexaF647- peptide nucleic acid (PNA) as the hybridization probes, because its electro-neutral polypeptide backbones enable high hybridization efficiency with the DNA templates [111, 112]. In addition, the dissociation rate of PNA/DNA hybrid is significantly slower than that of DNA duplexes, thus allowing the removal of non-hybridized, excess PNA from the reaction without reducing the fluorescence signals of the template-probe complexes due to dissociation of hybridized fluorescent probes. After fluorescent PNA probe hybridization and purification, the red and the green fluorescence intensities of the template-probe complexes are measured, and the resultant R/G fluorescence ratio is used to identify the DNA target (**Figure 3.1D**).

3.4 Ratiometric Fluorescence Detection with RCA

Using our RCA-based ratiometric fluorescence coding assay, we first demonstrate the detection of six infectious disease-related gene sequences with distinct fluorescence ratios. To do so, we used six synthetic oligonucleotides derived from conserved gene sequences of PM, HIV-1, NG, CT, TP and EC as the targets, challenged each of them to our mix of six padlock probes, and measured the red and the green fluorescence intensities of each sample at the end of the RCA reaction and probe hybridization. Fluorescent images of the six samples reveal that the six targets indeed yielded fluorescence ratios that were noticeably different from each other (**Figure 3.2A**). Moreover, the measured R/G ratio of each target closely matched its designed R/G ratio. For example, for the PM sample, the corresponding padlock probe of which is inserted with only green fluorescent probe binding sites, we observed only green fluorescence and no red fluorescence was detected, resulting in a measured R/G ratio of 0. While for the HIV-1 sample we observed weak red fluorescence and strong green fluorescence and measured a R/G ratio of ~0.5. In contrast, for the TP sample, we observed strong red fluorescence and weak green fluorescence, and measured a R/G ratio of ~3.0. Whereas for the EC sample, the padlock probe of which contains only red fluorescent probe sequence, we only observed fluorescence in the red channel and thus the R/G ratio was measured to be infinity. These fluorescence detection results clearly indicate that our RCA-based ratiometric fluorescence coding assay can achieve multiplexed detection of nucleic acid targets based on distinct R/G ratios.

Our RCA-based ratiometric assay detects the six infectious disease-related gene sequences with high reproducibility. Here, we replicated our assay for PM, HIV-1, NG, CT, TP and EC synthetic targets in four separate experiments, and plotted the average R/G ratios with error bars based on the four independent measurements (**Figure 3.2B**). The close match between the measured R/G ratios and the pre-designed R/G ratios provides strong support for the design principle of ratiometric fluorescence coding. In addition, the small error bars of measured ratios indicate that each target-specific ratio is indeed significantly different from each other and that our assay is highly reproducible. Taken together, these results also suggest that additional R/G ratios can be incorporated into our RCA-based assay by inserting additional binding sites in the padlock probes, which can expand the multiplexed detection capacity of the assay.

3.5 Ratiometric Fluorescence Detection with HRCA

Toward improving the sensitivity of multiplexed detection via our ratiometric fluorescence coding method, we show the compatibility between our method and HRCA. Compared with the linear amplification mechanism in RCA, HRCA employs a second primer to achieve exponential amplification of double-stranded DNA (dsDNA) templates (**Figure 3.3**) for fluorescent probes [113, 114]. Our PNA probe plays a critical role here, because strong binding affinity between PNA and DNA ensures our PNA probes can hybridize to the DNA templates even in the presence of complementary DNA strands. To verify the HRCA-based assay, we mixed our four padlock probes with each of the HIV-1, NR, CT and TP synthetic targets and detected the red and the green fluorescence intensities of each sample at the end of the HRCA reaction and probe hybridization. Similar to the results from RCA reactions, fluorescent images of these four HRCA samples also show fluorescence ratios that were noticeably distinct (**Figure 3.4A**). For example, we again observed relatively weak red fluorescence and strong green fluorescence for the HIV-1 sample, yielding a measured R/G ratio of ~0.8. On the other hand, strong red fluorescence and weak green fluorescence could be observed from the CT sample and the TP sample, with measured R/G ratios of ~3.4 and ~5.2, respectively. We similarly replicated the experiments four times and plotted the measured R/G ratios with errors bars (**Figure 3.4B**). The results indicated high reproducibility of our HRCA-based assay for detecting STD-related gene sequences based on distinct R/G fluorescence ratios.

Having verified robust multiplex detection through both our HRCA-based and RCA-based ratiometric fluorescence coding assays, we note that the measured R/G ratio for each HRCA sample was higher than that for its RCA counterpart. The difference in measured R/G ratios may be caused by the difference in self-quenching of neighboring Alexa Fluro 647 molecules between the HRCA and the RCA samples [115, 116]. Specifically, in RCA, the template-probe complex that is formed when thousands of probes hybridize onto long single-stranded DNA templates could fold into a spherical conformation in solution [117, 118]. This conformation confines the Alexa Fluro 647 molecules, and their proximity causes them to quench each other, thus reducing the red fluorescence intensity in RCA samples. In contrast, double strand DNA templates produced from HRCA were only hundreds of base pairs in length (**Figure 3.3**), which allowed much fewer fluorescent probes to hybridize onto each template and reduced the potential for self-quenching of neighboring Alexa Fluro 647 molecules. As a result, the intensities of red fluorescence in HRCA samples were elevated, resulting in higher R/G ratios.

3.6 Sensitivity of Ratiometric Fluorescence Coding Assay

Finally, we examine the sensitivity of our HRCA-based ratiometric fluorescence coding assay and its ability to detect genomic DNA. Here, we titrated our four STD synthetic targets to 1 pmol, 1 fmol, 1 amol to 1 zmol, tested them with our HRCA-based assay, and measured their red and green fluorescence intensities. For each target, as the input concentration lowered, both the red and the green fluorescence intensities expectedly reduced as a result of lower concentration of dsDNA templates generated from lower concentration of target, though all four targets were still detected at 1 zmol (**Figure 3.5A**). More importantly, for each target, even as the input concentration decreased by 9 orders of magnitude, the measured R/G ratios remained fairly consistent (**Figure 3.5B**). For example, the R/G rations for NR synthetic targets remained at ~2 for the 4 input target concentrations. The consistency of measured R/G ratios for the 4 targets across a wide range of input concentrations further illustrates the robustness of our ratiometric fluorescence coding method. Finally, we also used our HRCA-based assay to detect 200 ng and 2 ng NR genomic DNA (extracted from *Neisseria gonorrhoeae* Strain FA1090, ATCC). We observed strong red and green fluorescence in the sample with 200 ng NR genomic DNA sample and noticeably weaker fluorescence in the sample with 2 ng NR genomic DNA (**Figure 3.5C**, left). Moreover, the measured R/G ratios in both genomic DNA samples were comparable to that of the NR synthetic target (**Figure 3.5C**, right). These results therefore suggest that our HRCA-based ratiometric fluorescence coding assay can also detect STD genomic DNA based on its pre-designed fluorescence ratio.

3.7 Digital Ratiometric Fluorescence Coding

In previous chapters, we have demonstrated our Ratiometric Fluorescence Coding method for the detection of DNA targets from six infectious diseases using only two distinctly colored fluorescent probes However, the previous interaction of this approach is limited to identifying only one single pathogen among a number of suspected candidates, and its performance may be compromised by coinfections in clinical samples. Herein we have improved its utility for multiplexed nucleic acid detection by coupling Ratiometric Fluorescence Coding assay with a droplet-based microfluidic system. By encapsulating single copy of circularized padlock probe into droplet for amplification and measuring R/G fluorescence ratio of each droplet after fluorescent labeling, this approach allows simultaneously detecting multiple targets in a single reaction, termed Digital Ratiometric Fluorescence Coding.

Similar to our Ratiometric Fluorescence Coding assay described in **Chapter 3.3**, the Digital Ratiometric Fluorescence Coding also utilizes padlock probe chemistry coupled with HRCA to transform each target into a specific DNA template that allows hybridization of red and green molecular beacons (MBs) at a pre-designed ratio in single droplet (**Figure 3.6A**). Similar to previous work, we design each padlock probe to contain a target-specific hybridization region, a unique ratio of binding sites for red and green MBs and general primer sites. In the presence of a DNA target, the corresponding linear padlock probe is ligated to form a circular probe, which is then capsulated into a ~3pL droplet for subsequent HRCA. Like digital PCR […], the circular probes are distributed into these droplets following the Poisson distribution, where there are some droplets that have no circular probe, and others that have one or more copies of circular padlock probe encapsulated. In our work, the mixture for amplification reaction were appropriately diluted to make sure that most of droplets encapsulate either zero or single copy of circular probe. A PDMS microfluidic device (**Figure 3.7**) is fabricated for droplet generation and measurement. Droplets are generated through a 10 μm flow-focusing junction on the device and then incubated at 60 °C for 2 hours. Within those droplets containing circular probes, HRCA amplifies along the circular probe to generate DNA templates with hundreds of tandem repeats that serve as hybridization sites for MBs. After MB hybridization, red and green fluorescence intensities of each droplet are measured through a custom confocal microscope system (**Figure 3.8**), and the resultant red to green (R/G) fluorescence ratio is used to identify the DNA target. After MB hybridization, droplets were observed using a fluorescence microscope under white light, red channel and green channel (**Figure 3.6B**). The droplets presenting fluorescence signals indicated that a single copy of circular padlock probe was capsulated into the droplet and amplified via HRCA, while no fluorescence signal was observed in the droplets without circular probes.

For a sample with multiple targets, the hybridization and ligation reaction generate multiple circular probes with distinct pre-designed MB ratios encode in initially linear padlock probes. After droplet generation, single copy of circular probes with distinct pre-designed MB ratios are encapsulated into different droplets. After amplification reaction and MB hybridization, theses droplets containing different circular probes, from which DNA templates with distinct hybridization sites for MBs are generated via HRCA, present distinct R/G fluorescence colors that can be used to identify the DNA targets in the initial sample. Using our Digital Ratiometric Fluorescence Coding method, we demonstrate the multiplexed detection of gene sequences from up to 6 STDs, with distinct and highly resolved fluorescence ratios, indicating that our method offers ready means for expanding the multiplexed detection capacity of NAATs.

3.8 Digital Ratiometric Fluorescence Detection of Single Target

In this work, we designed six circularizable padlock probes to detect 6 common STDs. The target-specific regions of these 6 padlock probes target conserved sequences from *Microplasma genitalium* (MG), *Herpes simplex virus* (HSV), *Neisseria gonorrhoeae* (NR), *Human immunodeficiency virus* (HIV), *Treponema pallidum* (TP) and *Chlamydia trachomatis* (CT). Moreover, we incorporated the binding sites for the red and the green fluorescent probes at specific red to green (R/G) ratios into these padlock probes: 0:1 for MG, 1:4 for HSV, 1:1 for NG, 2:1 for HIV-1, 4:1 for TP and 1:0 for CT, respectively. Using our Digital Ratiometric Fluorescence Coding method, we demonstrate the multiplexed detection of gene sequences from the six STDs with distinct fluorescence ratios.

Using our Digital Ratiometric Fluorescence Coding method, we first demonstrate the detection of six STD-related gene sequences with distinct fluorescence ratios. To do so, we used six synthetic oligonucleotides derived from conserved gene sequences of MG, HSV, NG, HIV, TP and CT as the targets, challenged each of them to our mix of six padlock probes, and measured the red and the green fluorescence intensities of droplets from each sample. After fluorescence measurement on the confocal microscope system, samples with different DNA targets indeed yielded R/G fluorescence ratios that were noticeably distinct from each other (**Figure 3.9**). For example, for the MG sample, the corresponding padlock probe of which is inserted with only green MB binding sites, we observed only green fluorescence and no red fluorescence was detected. While for the NG sample with pre-designed R/G fluorescence ratio of 1:1, the detected peak heights of red fluorescence in the droplets are close to those of green fluorescence. In contrast, for the HIV-1 sample, the padlock probe of which is designed to have R/G fluorescence ratio of 2:1, we observed strong red fluorescence and weak green fluorescence in the measured droplets, while for the TP sample with pre-designed R/G ratio of 4:1, even stronger red fluorescence signal and weaker green fluorescence signal were observed. Whereas for the CT sample, the padlock probe of which contains only red MB sequence, we only observed fluorescence signal in the red channel and no green fluorescence signal. These fluorescence detection results clearly indicate that our Digital Ratiometric Fluorescence Coding assay can identify the nucleic acid targets based on distinct R/G ratios. Moreover, the scatter plot for red and green fluorescence intensities in each sample revealed that the six target-specific R/G fluorescence ratios can be unambiguously separated (**Figure 3.10**).

We also calculated average measured R/G fluorescence ratios and standard deviation of R/G ratios from over thousands of droplets in each sample (**Figure 3.11A**). The results also showed that six measured fluorescence ratios were noticeably distinct and indicated our Digital Ratiometric Fluorescence Coding assay for detecting STD-related gene sequences based on distinct R/G fluorescence ratios. Moreover, the strong linearity between the measured R/G ratios and the pre-designed R/G ratios (slope= 1.6643 and R^2 = 0.9973, **Figure 3.11B**) provides strong support for the design principle of our Ratiometric Fluorescence Coding. We note that the average measured R/G fluorescence ratio for each sample was higher than that for its pre-designed R/G ratios. Similar to the discussion about how self-quenching of neighboring Alexa Fluro 647 molecules in **Chapter 3.5**, though the R/G ratios are different, we think it shouldn't affect the demonstration of our Ratiometric Fluorescence Coding concept. Based on the experimental protocol of our assay, there are six key factors that affects the measured R/G ratios: 1) pre-designed ratios of red and green fluorescent probe/MB sequences on padlock probe, 2) fluorescence quantum yield, 3) hybridization efficiency of fluorescent probe/MB, 4) fluorophore self-quenching, 5) excitation laser power, and 6) fluorescence measurement. As we know, pre-designed R/G ratio on padlock is the main factor that leads to distinct measured R/G ratios. We also have known that the fluorescence quantum yields for these two dyes, AlexaF647 and AlexaF488, are 0.33 and 0.92 respectively, and two laser powers on the custom confocal microscope system are not the same in our experiments. However, there are three other main factors, probe hybridization efficiency, fluorophore self-quenching and instrument detection efficiency, which are not identical for these two dyes, but cannot be quantified in the experiments. The existence of other unmeasurable factors, together with known or measurable factors, may result in the difference between measured R/G fluorescence ratios and pre-designed R/G ratios. Moreover, the essence of our Ratiometric Fluorescence Coding concept is to encode each nucleic acid target with a specific fluorescence ratio, thus the measured R/G ratios are not necessary to be the same as pre-designed ratios once the measured ratios are clearly differentiable and robustly reproducible for the identification of each nucleic acid target, which has been demonstrated using six STD target sequences.

3.9 Digital Ratiometric Fluorescence Detection of Multiple Targets Simultaneously

We have demonstrated our Digital Ratiometric Fluorescence Coding method for multiplexed detection of different STD-related gene sequences in a single reaction. First, we mixed both NG and HIV synthetic targets with our six padlock probes for hybridization and ligation in a single reaction. After droplet generation and amplification, we measured the red and green fluorescence intensities of each droplet. The resultant scatter plot for red and green fluorescence intensities present two populations that are unambiguously separated (**Figure 3.12A**). Moreover, the average measured R/G ratios of NG population (**Figure 3.12A, yellow dots**) is around 1.6, close to the measured R/G ratio resultant from the sample with NG target only (**Figure 3.11A**), while the average measured R/G ratios of HIV population (**Figure 3.12A, yellow dots**) is ~3.2, close to the R/G ratio of single target HIV sample (**Figure 3.11A**).

In addition, we mixed four different STD synthetic targets, MG, CT, NG and HIV, in a single reaction and tested using our Digital Ratiometric Fluorescence Coding assay. In the resultant scatter plot for red and green fluorescence intensities measured, four distinct population are clearly separated (**Figure 3.12B**). More specifically, there is one population of droplets with strong green fluorescence signals and nearly no red fluorescence signals, which demonstrates the presence of MG targets, the padlock probe of which is inserted with only green MB binding sites, we observed only green fluorescence and no red fluorescence was detected. In contrast, another droplet population with no green fluorescence and strong red fluorescence represents the CT target, of which the padlock probe has only red MB binding sites. The other two populations look the same as the two distinct populations we observed in the sample with the mixture of NG and HIV targets (**Figure 3.12A**). For further demonstration, we mixed all six STD targets and the resultant scatter plot for red and green fluorescence intensities presents six well-separated populations (**Figure 3.12C**). Besides the four populations observed in **Figure 3.12B**, there is one population showing up between the two populations presenting MG and NG (**Figure 3.12C, cyan**), which demonstrates the presence of HSV target with pre-designed R/G ratio of 1:4. While another additional population between HIV and CT represents the existence of TP, of which the padlock probe has pre-designed ratio of 4:1 (**Figure 3.12C, magenta**). These results have demonstrated our Digital Ratiometric Fluorescence Coding method for multiplexed detection of nucleic acid targets based on distinct R/G fluorescence ratios.

3.9 Conclusion

In this chapter, we have developed a new ratiometric fluorescence coding concept that can be employed to significantly expand multiplexing capacity of NAAT by assigning a specific fluorescence color ratio for coding each target. Unlike common one-color-one-target scheme, we produce a target-specific DNA template with a pre-designed ratio of binding sites for distinctly colored fluorescent probes. Our concept has been validated using padlock probe chemistry coupled with RCA or HRCA reactions. With our custom-designed padlock probes targeting infectious disease-related genes, we have demonstrated our ratiometric fluorescence coding method for multiplexed detection of 6 nucleic acid targets using only two distinctly colored fluorescent probes. In addition, our HRCA-based assay has been utilized to detect genomic DNA from a specific STD.

Furthermore, we have coupled the assay with a droplet-based microfluidic system to extend its utility to simultaneously detecting multiple targets in a single reaction. With our custom-designed padlock probes targeting STD-related genes, we have demonstrated our Digital Ratiometric Fluorescence Coding method for multiplexed detection of 6 nucleic acid targets using only two distinctly colored molecular beacons, as well as its potential for further expanding the multiplexing capacity.

There are several areas to improve our method for highly multiplex and sensitive detection of nucleic acids. First, the multiplexing capacity of our method can be readily expanded by using more fluorescence colors and designing more fluorescent probe binding sites. For example, using 3 colors and 5 binding sites (S=5), the maximum number of ratio can reach 40 (**Table 3.3**). Since the commercially available phosphorylated single-stranded DNA is limited to 200 bp from Integrated DNA Technologies, Inc. (Coralville, IA), decreasing the length of fluorescent probes can be an alternative approach to provide more available probe binding sites on a padlock probe. Especially using high affinity DNA analogues, such as PNA, as fluorescent probes, the probe length can be decreased to 12-mer or even shorter [119, 120], which will allow more than 10 binding sites on a single padlock probe. In addition, we can also perform multiplexed detection of nucleic acid by coupling our Ratiometric Fluorescence Coding concept with single molecule detection method [121, 122]. The molecule-by-molecule measurement scheme in single molecule detection system can measure fluorescently-labeled template-probe complexes one at a time in both red and green channel simultaneously, thus allows differentiation of template-probe complexes from different targets based on their distinct R/G ratios, which can be utilized for simultaneous detection of multiple targets in a single reaction. In addition to RCA and HRCA reactions, we also envision that strategies for coupling our Ratiometric Fluorescence Coding concept with other nucleic acid amplification techniques such as LCR and PCR can be readily developed. Therefore, given the good performance of our ratiometric fluorescence coding method and the means to improve it, we believe our method has the potential to perform large-scale multiplexed detection of nucleic acids upon further development.

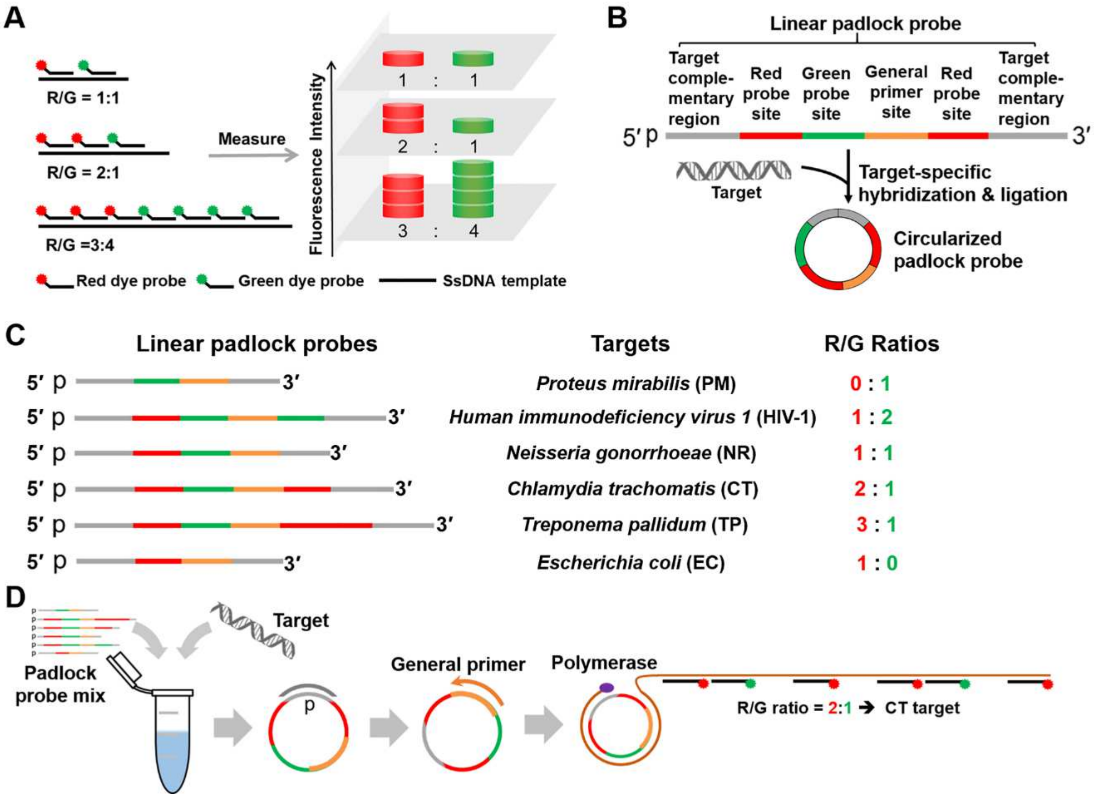

A
R/G = 1:1
R/G = 2:1
R/G =3:4
Measure
Fluorescence Intensity
1 : 1
2 : 1
3 : 4
Red dye probe
Green dye probe
SsDNA template
B
Linear padlock probe
Target comple-mentary region
Red probe site
Green probe site
General primer site
Red probe site
Target comple-mentary region
5′ p
3′
Target
Target-specific hybridization & ligation
Circularized padlock probe
C
Linear padlock probes
Targets
R/G Ratios
5′ p 3′
Proteus mirabilis (PM)
0 : 1
Human immunodeficiency virus 1 (HIV-1)
1 : 2
Neisseria gonorrhoeae (NR)
1 : 1
Chlamydia trachomatis (CT)
2 : 1
Treponema pallidum (TP)
3 : 1
Escherichia coli (EC)
1 : 0
D
Padlock probe mix
Target
p
General primer
Polymerase
R/G ratio = 2:1 → CT target

Figure 3.1. Overview of Ratiomatric Fluorescence Coding for multiplexed nucleic acid detection

(A) Our ratiometric fluorescence coding concept utilizes a combination of fluorophore colors and their fluorescence intensity levels to achieve multiplexed detection. For example, upon hybridization of fluorescent probes with two colors (red and green) onto a single DNA template with a specific ratio, the template-probe complex can be decoded by resolving the intensity level of each color and display the specific red to green (R/G) fluorescence ratio. **(B)** The internal part of our padlock probe consists of one or two general primer sequences common for all the padlock probes and several red and green fluorescent probe sequences, which have a unique R/G ratio for each padlock probe. In the presence of corresponding target, padlock probe is ligated to a closed circle for subsequent amplification. **(C)** Six padlock probes with different R/G ratios were designed for infectious disease detection. Six circularizable padlock probes encoded as R/G ratio of 0:1, 1:2, 1:1, 2:1, 3:1 and 1:0 are designed for targeting gene sequences in *Proteus mirabilis* (PM), *Human immunodeficiency virus 1* (HIV-1), *Neisseria gonorrhoeae* (NR), *Chlamydia trachomatis* (CT), *Treponema pallidum* (TP), and *Escherichia coli* (EC) respectively. **(D)** RCA-based ratiometric fluorescence assay starts with the mixture of our six target-specific padlock probes and infectious disease-related DNA target. When binding to a specific target, the phosphorylated 5' end and the 3' end of the padlock probe are brought into proximity that leads to formation of a circularized probe via ligation. Subsequently, RCA reaction is performed to generate long ssDNA with thousands of tandem repeats complementary to the padlock probe sequences, serving as binding sites for fluorophores-labeled PNA probes. Upon hybridization with PNA-AF647 (red) and PNA-AF488 (green), the fluorophores-labeled amplification products are scanned in both green and red channels and the measured R/G fluorescence ratios can be used to identify DNA targets in the initial input. Adapted with permission from [123]. Copyright 2018 American Chemical Society.

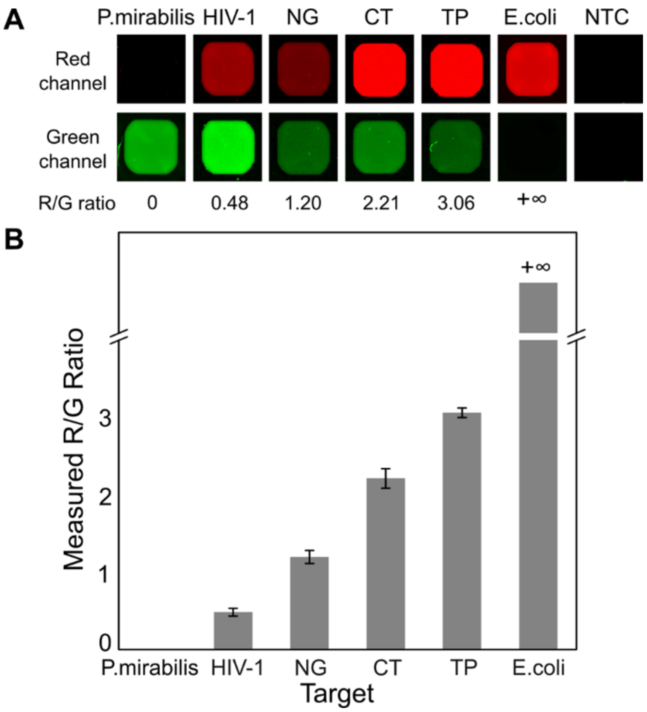

A
P.mirabilis
HIV-1
NG
CT
TP
E.coli
NTC
Red channel
Green channel
R/G ratio
0
0.48
1.20
2.21
3.06
+∞
B
+∞
Measured R/G Ratio
0
1
2
3
P.mirabilis
HIV-1
NG
CT
TP
E.coli
Target

Figure 3.2. RCA-based ratiometric fluorescence detection of six gene sequences

(A) RCA-based ratiometric fluorescence coding assay was validated via detecting six infectious disease-related genes. With PM, HIV-1, NR, CT, TP, EC synthetic targets and H_2O control, the resultant fluorophores-labeled RCA products were scanned under both red (top row) and green (bottom row) channels. From left to right in each channel, the fluorescent spots represent fluorophores-labeled RCA products resultant from PM, HIV-1, NR, CT, TP, EC synthetic targets and H_2O control, respectively. Moreover, measured R/G ratio for each sample is close to the pre-designed ratio. The spots of fluorophores-labeled PM amplification products present fluorescence only in green channel with a measured R/G ratio of 0. The spots of HIV-1 sample show weak fluorescence in red channel and strong fluorescence in green channel with a measured R/G ratio of 0.48. The spots of NG sample show nearly equivalent fluorescence intensities in red and green channel with a measured R/G ratio of 1.20. In contrast, the spots of CT and TP samples show strong fluorescence in red channel and weak fluorescence in green channel with measured R/G ratios of 2.21 and 3.06 respectively. And the spots of EC sample show fluorescence only in the red channel with a measured R/G ratio of ∞. The spot of H_2O control sample shows no fluorescence signal in either red or green channel. **(B)** Our assay was replicated in four independent experiments for detecting the six infectious disease-related synthetic targets. The small error bars of measured R/G ratios indicate a robust reproducibility. Adapted with permission from [123]. Copyright 2018 American Chemical Society.

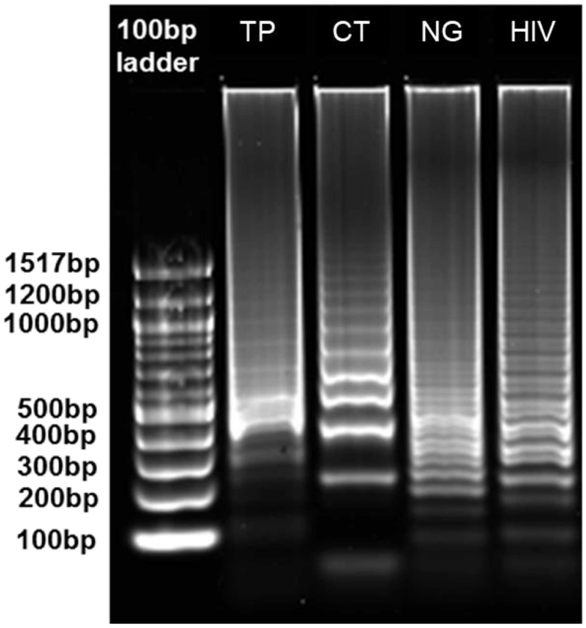

100bp ladder
TP
CT
NG
HIV
1517bp
1200bp
1000bp
500bp
400bp
300bp
200bp
100bp

Figure 3.3. Gel image of HRCA products

Agarose Gel electrophoresis result of HRCA products of TP, CT, NG and HIV sample. Adapted with permission from [123]. Copyright 2018 American Chemical Society.

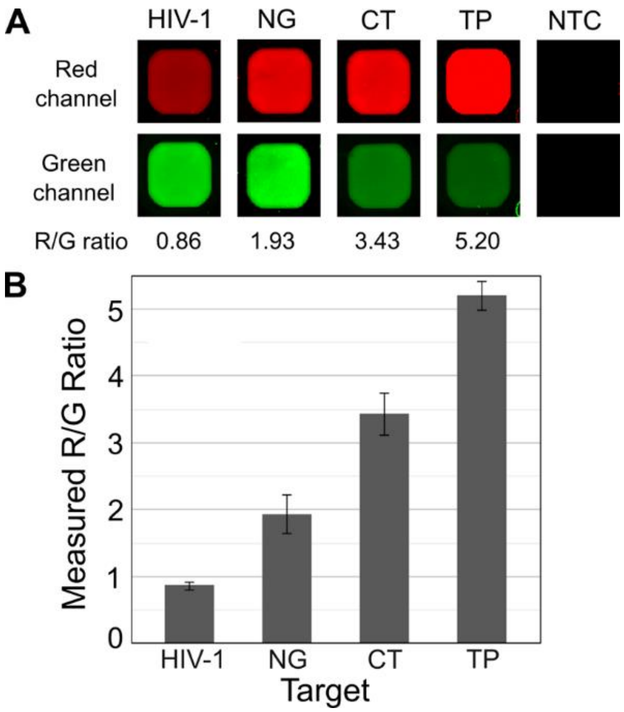

A
HIV-1
NG
CT
TP
NTC
Red channel
Green channel
R/G ratio
0.86
1.93
3.43
5.20
B
Measured R/G Ratio
0
1
2
3
4
5
HIV-1
NG
CT
TP
Target

Figure 3.4. HRCA-based ratiometric fluorescence detection of four gene sequences

(A) HRCA-based ratiometric fluorescence coding assay was verified using our four padlock probes and STD synthetic targets. In the presence of HIV-1, NR, CT and TP synthetic targets, HRCA generates dsDNA amplification products that are hybridized with red and green fluorescent PNA probes. The resultant fluorophores-labeled HRCA products were scanned under Typhoon image scanner in both red (top row) and green (bottom row) channels. From left to right in each channel, the fluorescent spots represent fluorophores-labeled HRCA products of HIV-1, NR, CT, TP and H_2O control samples respectively. The spots of fluorophores-labelled HIV-1 amplification products show weak fluorescence in red channel and strong fluorescence in green channel with measured R/G ratios of 0.86. While the spots of NG, CT and TP samples show strong fluorescence in red channel and weak fluorescence in green channel with measured R/G ratios of 1.93, 3.43 and 5.20 respectively. The spot of H_2O control sample shows no fluorescence signal in either red or green channel. **(B)** Our HRCA-based assay was also replicated in four separated experiments. Each measured target-specific R/G ratio is significantly different from each other and their small error bars indicate high reproducibility of our HRCA-based assay. Adapted with permission from [123]. Copyright 2018 American Chemical Society.

.

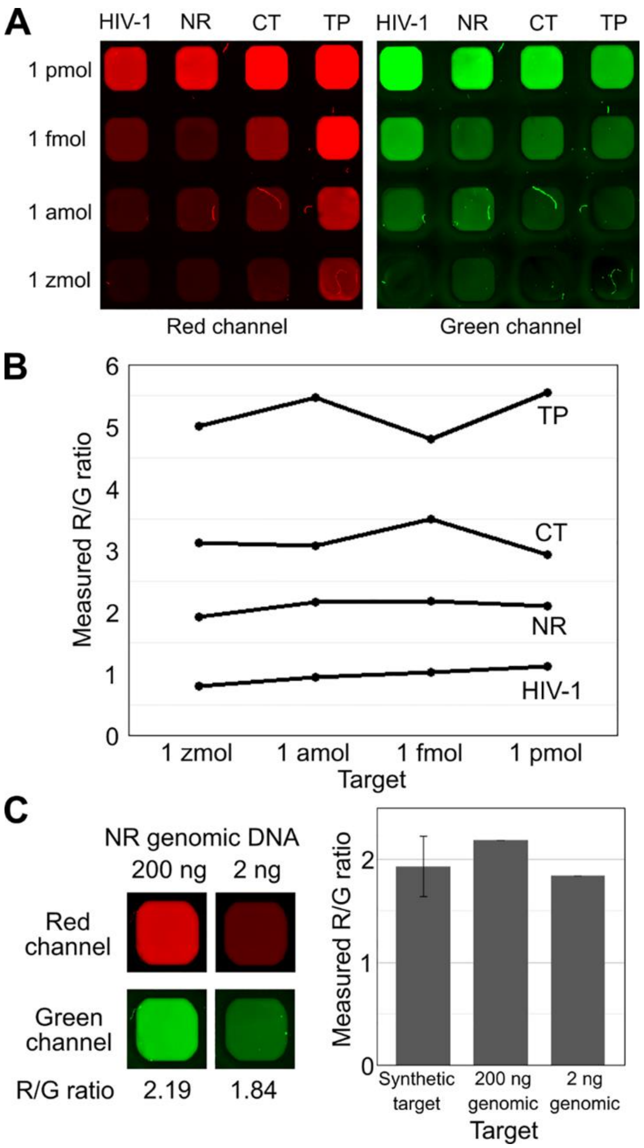

A
HIV-1
NR
CT
TP
HIV-1
NR
CT
TP
1 pmol
1 fmol
1 amol
1 zmol
Red channel
Green channel
B
Measured R/G ratio
0
1
2
3
4
5
6
TP
CT
NR
HIV-1
1 zmol
1 amol
1 fmol
1 pmol
Target
C
NR genomic DNA
200 ng
2 ng
Red channel
Green channel
R/G ratio
2.19
1.84
Measured R/G ratio
0
1
2
Synthetic target
200 ng genomic
2 ng genomic
Target

Figure 3.5. Sensitivity of HRCA-based ratiometric fluorescence coding

(A) The sensitivity of our HRCA-based ratiometric fluorescence coding method was verified using a gradient input of each STD synthetic targets ranging from 1 pmol, 1 fmol, 1 amol to 1 zmol. The fluorophore-labeled HRCA samples were scanned under Typhoon scanner in both red (left panel) and green (right panel) channels. The synthetic target inputs are HIV-1, NR, CT and TP from left to right in each row, and target input concentrations are 1 pmol, 1 fmol, 1 amol and 1 zmol from top to bottom in each column. The fluorescence intensities in both red and green channels were observed to decline as target input decreasing due to the deduction of dsDNA templates generated through HRCA. **(B)** For each target, the measured R/G ratios remained fairly consistent across the input concentrations, further indicating the robustness of our method. **(C)** Our HRCA-based assay was performed to detect 200 ng and 2ng of NR genomic DNA. The resultant fluorophores-labeled HRCA samples were scanned under in both red and green channels, and R/G fluorescence ratios were measured (left). R/G ratios of two NR genomic DNA samples were measured to be comparable to that of NR synthetic target using our HRCA-based assay (right). Adapted with permission from [123]. Copyright 2018 American Chemical Society.

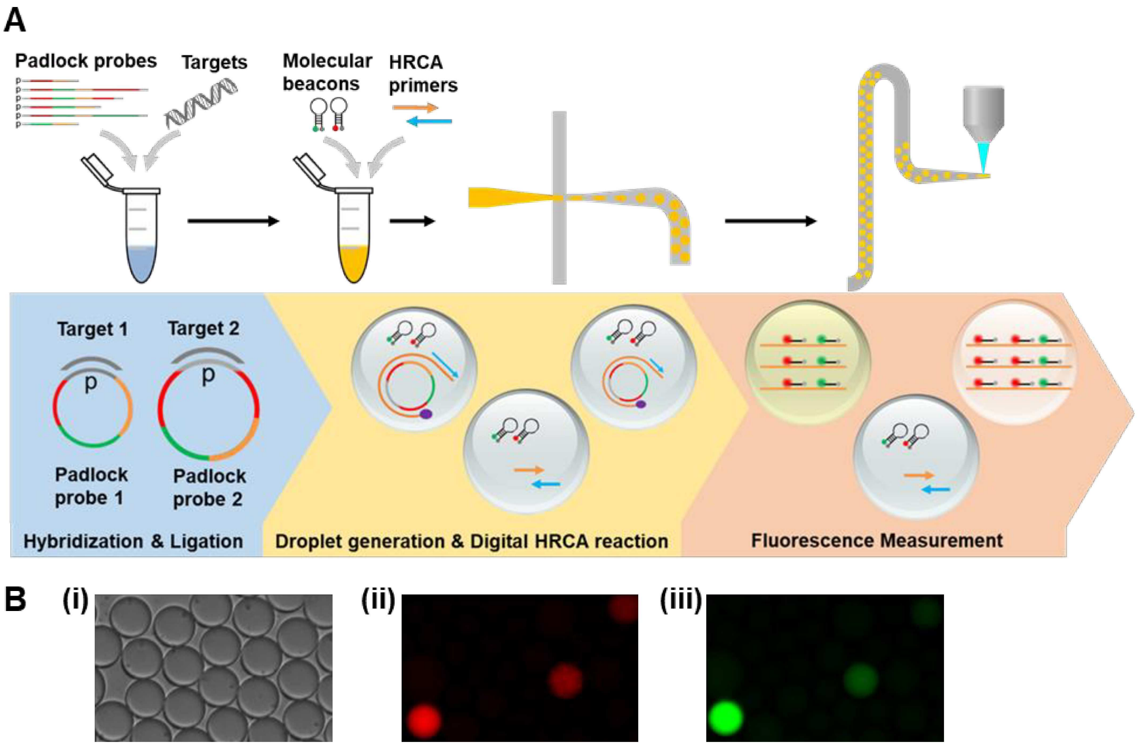

A
Padlock probes
Targets
Molecular beacons
HRCA primers
Target 1
Target 2
p
p
Padlock probe 1
Padlock probe 2
Hybridization & Ligation
Droplet generation & Digital HRCA reaction
Fluorescence Measurement
B
(i)
(ii)
(iii)

Figure 3.6. Schematic illustration of Digital Ratiometric Fluorescence Coding method

(A) We start with the mixture of our 6 target-specific padlock probes and DNA targets. After ligation, HRCA and hybridization of AF555- (red) and AF488- (green) labeled MBs, the droplets are measured on a custom confocal microscope system**. (B)** Droplets images. After MB hybridization, droplets were observed using a fluorescence microscope under white light (i), red channel (ii) and green channel (iii). Droplets with fluorescence indicated that at least one copy of circular padlock probe was capsulated into the droplet, while droplets without fluorescence suggesting no circular probes.

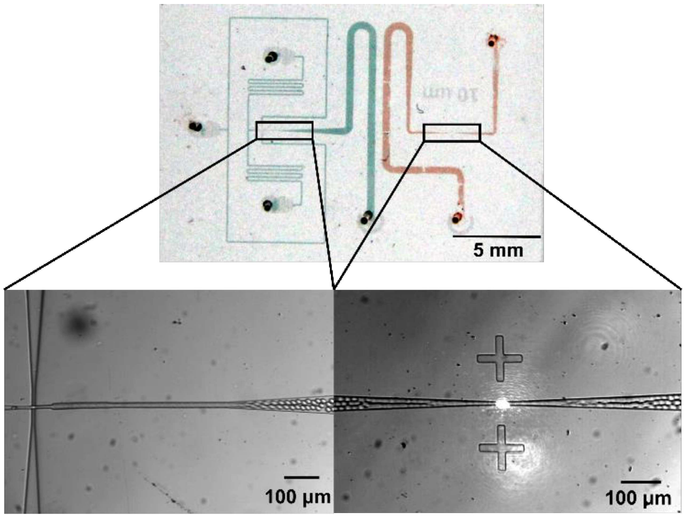

10 mm
5 mm
100 µm
100 µm

Figure 3.7. Image of Microfluidic device for droplet generation and measurement

The microfluidic device consists of a droplet generation module (green) and a droplet measurement module (red). Droplets are generated via a 10 µm flow-focusing junction (bottom left) and measured when flowing through a 10 µm constriction channel (bottom right).

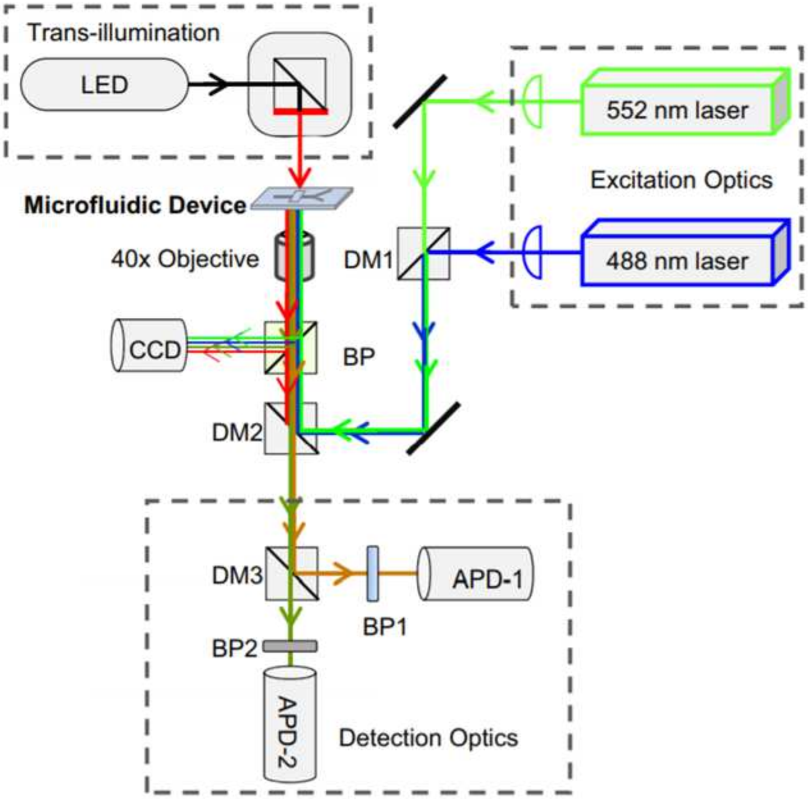

Trans-illumination
LED
552 nm laser
Excitation Optics
488 nm laser
Microfluidic Device
40x Objective
DM1
CCD
BP
DM2
DM3
APD-1
BP1
BP2
APD-2
Detection Optics

Figure 3.8. A schematic of the custom confocal microscope system

The optical setup consists of dual laser excitation (488nm and 552nm) and dual band fluorescence detection (506-534nm and 608-648nm) with two silicon avalanche photodiode detectors (APD).

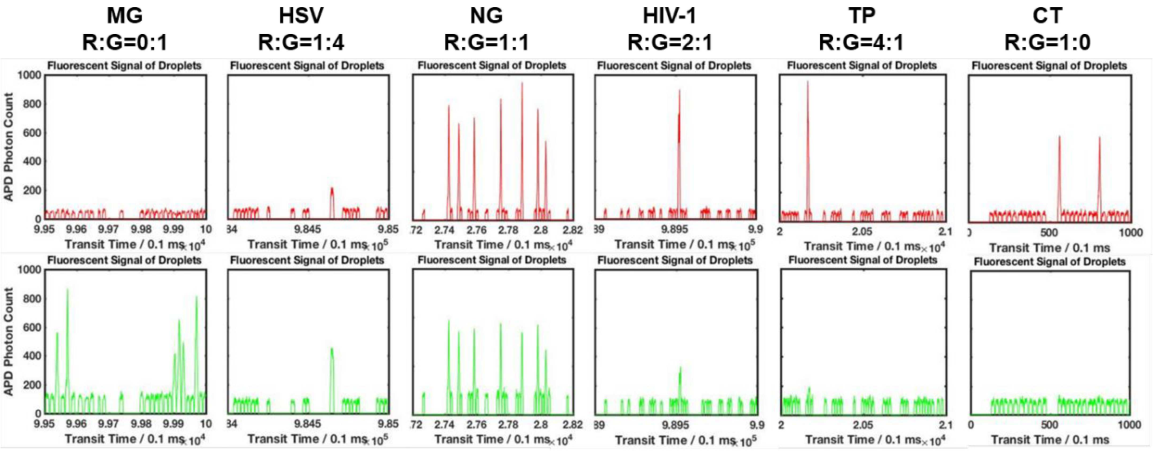

MG
R:G=0:1
HSV
R:G=1:4
NG
R:G=1:1
HIV-1
R:G=2:1
TP
R:G=4:1
CT
R:G=1:0
Fluorescent Signal of Droplets
APD Photon Count
0
200
400
600
800
1000
Transit Time / 0.1 ms
×10^4
×10^5
9.95
9.96
9.97
9.98
9.99
10
9.845
9.85
2.74
2.76
2.78
2.8
2.82
9.895
9.9
2
2.05
2.1
500
1000

Figure 3.9. Droplet fluorescence traces for different DNA samples

Six synthetic oligonucleotides derived from conserved gene sequences of MG, HSV, NG, HIV, TP and CT were used as the targets in the Digital Ratiometric Fluorescence Coding assay. The fluorescence measurement on the confocal microscope system revealed that samples with different DNA targets indeed yielded R/G fluorescence ratios that were noticeably distinct from each other. For example, for the MG sample with pre-designed R/G ratio of 0:1, only green fluorescence was detected and no red fluorescence observed, while for the CT sample with pre-designed R/G ratio of 1:0, only red fluorescence was detected and no green fluorescence was observed.

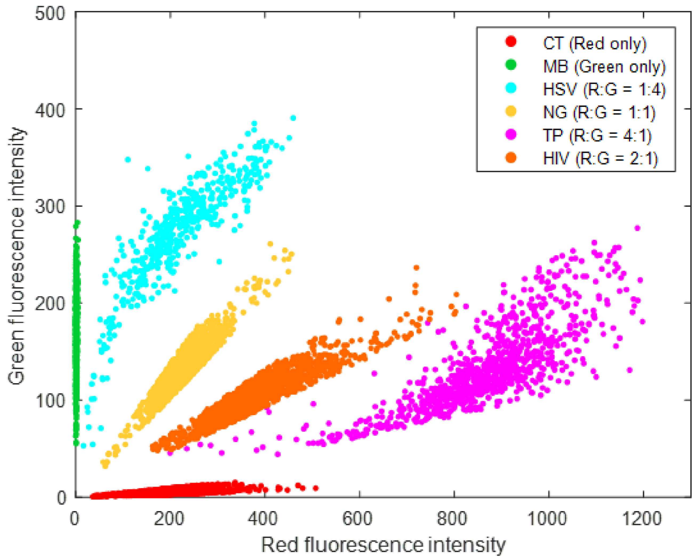

CT (Red only)
MB (Green only)
HSV (R:G = 1:4)
NG (R:G = 1:1)
TP (R:G = 4:1)
HIV (R:G = 2:1)
Green fluorescence intensity
Red fluorescence intensity
0
100
200
300
400
500
0
200
400
600
800
1000
1200

Figure 3.10. Comparison of single molecule counting and bulk fluorescence analysis

Scatter plot for red and green fluorescence intensities of six samples with distinct R/G ratios. We used xix synthetic oligonucleotides derived from conserved gene sequences of MG, HSV, NG, HIV, TP and CT as the targets, challenged each of them to our mix of six padlock probes, and measured the red and the green fluorescence intensities of droplets from each sample. Red and green fluorescence of each droplet were measured on the confocal microscope system and the scatter plot for red and green fluorescence intensities in each sample revealed that the six target-specific R/G fluorescence ratios were unambiguously separated.

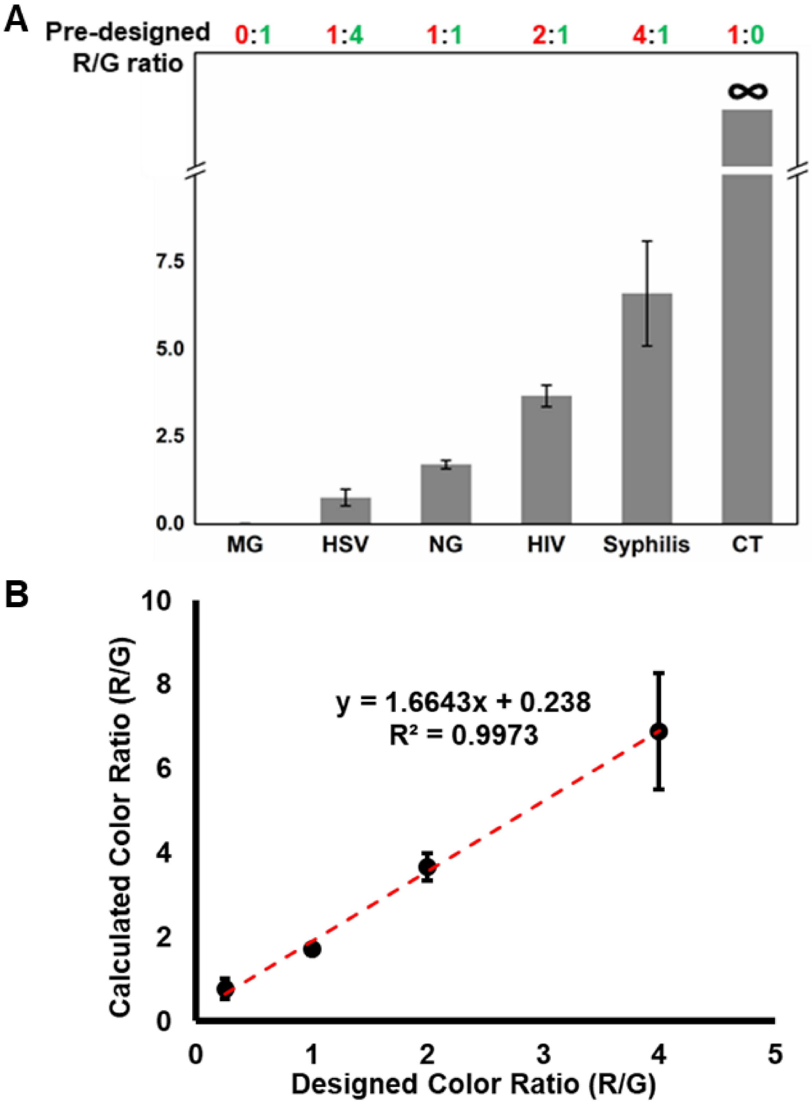

A
Pre-designed R/G ratio
0:1
1:4
1:1
2:1
4:1
1:0
∞
0.0
2.5
5.0
7.5
MG
HSV
NG
HIV
Syphilis
CT
B
y = 1.6643x + 0.238
R² = 0.9973
Calculated Color Ratio (R/G)
Designed Color Ratio (R/G)
0
2
4
6
8
10
0
1
2
3
4
5

Figure 3.11. Average R/G fluorescence ratios and their standard deviations

(A) Average R/G fluorescence ratios and their standard deviations were calculated from ratios measured on over thousands of droplets in each sample. The results have shown that six measured fluorescence ratios are noticeably distinct, indicating our Digital Ratiometric Fluorescence Coding assay for identifying STD-related gene sequences based on distinct R/G fluorescence ratios. (B) A linear regression is performed between average measured R/G ratios and pre-designed ones. The strong linearity between them with slope= 1.6643 and $R^2 = 0.9973$ also provides strong support for the design principle of our Ratiometric Fluorescence Coding.

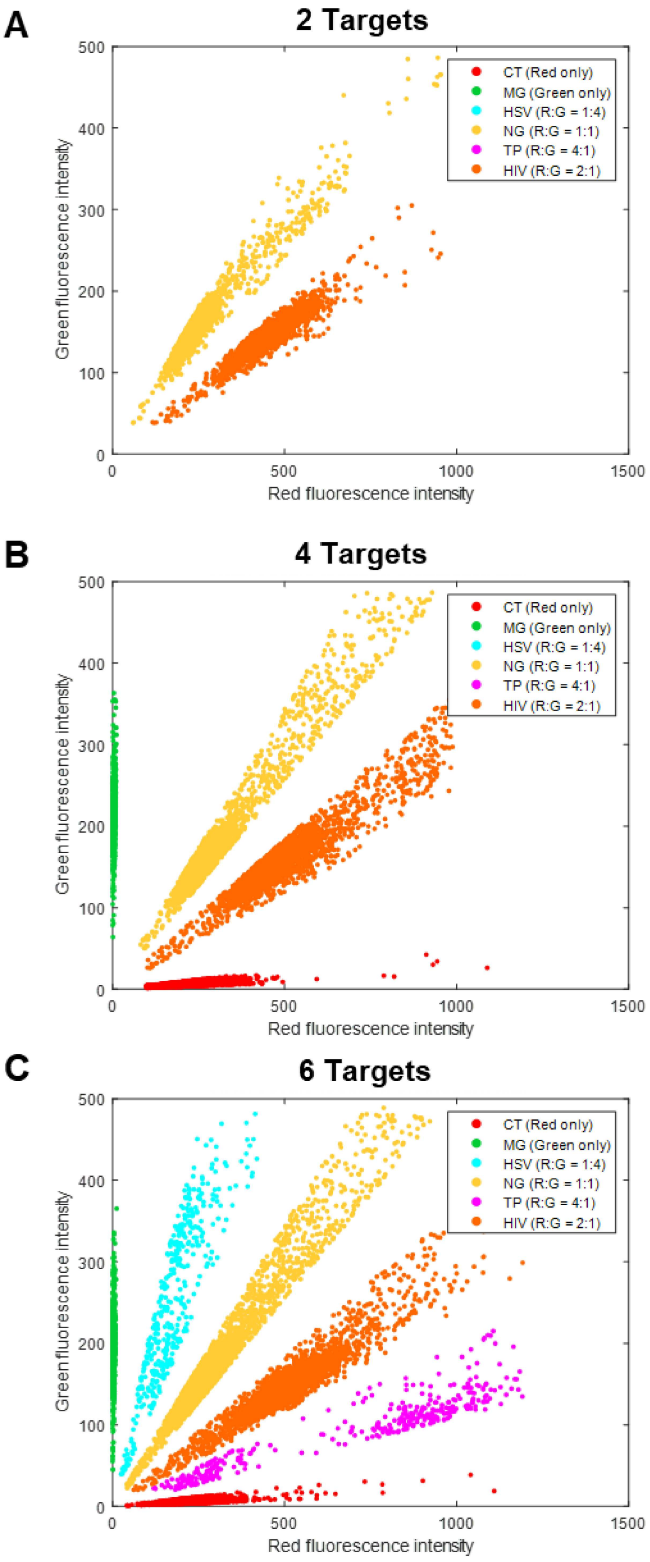

A
2 Targets
B
4 Targets
C
6 Targets
● CT (Red only)
● MG (Green only)
● HSV (R:G = 1:4)
● NG (R:G = 1:1)
● TP (R:G = 4:1)
● HIV (R:G = 2:1)
Green fluorescence intensity
Red fluorescence intensity
0
100
200
300
400
500
0
500
1000
1500

Figure 3.12. Digital Ratiometric Fluorescence Coding method for multiplexed detection of STD-related gene sequences in a single reaction

(A) Both NG and HIV synthetic targets were pre-mixed and tested using our Digital Ratiometric Fluorescence Coding. The resultant scatter plot for red and green fluorescence intensities present two populations that are unambiguously separated, indicating our assay for multiplexed detection of more than one target in a single reaction. Moreover, our Digital Ratiometric Fluorescence Coding has been demonstrated for multiplexed detection of four different STD-related gene sequences (B) and up to six different STD-related gene sequences (C) in a single reaction.

Padlock probes	**Sequence**
Padlock_PM	/5Phos/GACCAGTTTTGCTTGCGCCCCCTAACCCTAACCCTAAGACTCGTCATGTCTCAGCTTTAGCCCTAACCCTAACCCTAACGTGCTGCACGTTTAGCAC
Padlock_TP	/5Phos/CTGTCGCACCGTGAGTTCATCTCAGAACTCACCTGTTAGGACTCGTCATGTCTCAGCTTTAGCCCTAACCCTAACCCTAATTTTTCAGAACTCACCTGTTAGTTTTTCAGAACTCACCTGTTAGAAAGGGCTGATGGCTCTGAG
Padlock_CT	/5Phos/CCTTATGATCGACGGAATTCTGTGTCAGAACTCACCTGTTAGGACTCGTCATGTCTCAGCTTTAGCCCTAACCCTAACCCTAATTTTTCAGAACTCACCTGTTAGGGGAATCCTGCTGAACCAAG
Padlock_NG	/5Phos/GGAACCCGATATAATCCGCCCTTTCAGAACTCACCTGTTAGGACTCGTCATGTCTCAGCTTTAGCCCTAACCCTAACCCTAAATATTGTGTTGAAACACCGCCC
Padlock_HIV	/5Phos/CCAAATGAGAGAACCAAGGGGAAGGACTCGTCATGTCTCAGCTTTAGCCCTAACCCTAACCCTAATTTTTCAGAACTCACCTGTTAGTTTTCCCTAACCCTAACCCTAACAGGGCCTATTGCACCAGG
Padlock_EC	/5Phos/GGCGTGGTGTAGAGCATTACGTCAGAACTCACCTGTTAGGACTCGTCATGTCTCAGCTTTAGTCAGAACTCACCTGTTAGATATCGTCCACCCAGGTGTTC

Synthetic targets	**Sequence**
Target_PM	5’-GGCGCAAGCAAAACTGGTCGTGCTAAACGTGCAGCACG
Target_TP	5’-TAGTCGATGAACTCACGGTGCGACAGCTCAGAGCCATCAGCCCTTTTCAGC
Target_CT	5’-CACAGAATTCCGTCGATCATAAGGCTTGGTTCAGCAGGATTCCCCACAG
Target_NG	5’-AAGGGCGGATTATATCGGGTTCCGGGCGGTGTTTCAACACAATATGGCGG
Target_HIV	5’-TGTCACTTCCCCTTGGTTCTCTCATTTGGCCTGGTGCAATAGGCCCTGCATGC
Target_EC	5’-CG TAATGCTCTA CACCACGCCG AACACCTGGG TGGACGATAT

Table 3.1. Padlock probe and synthetic target sequences for Ratiometric Fluorescence Coding assay demonstration

Padlock probes targeting gene sequences in *Proteus mirabilis* (PM, 238:275 from AF240693) *Human immunodeficiency virus 1* (HIV-1, 258:305 from KC966998), *Neisseria gonorrhoeae* (NG, 362:411 from X52364), *Chlamydia trachomatis* (CT, 50:98 from JX648604), *Treponema pallidum* (TP, 801:846 from KC966998), and *Escherichia coli* (EC, 189:230 from AY447194) were designed for the demonstration of Ratiometric Fluorescence Coding concept. Adapted with permission from [123]. Copyright 2018 American Chemical Society.

Padlock probes	Sequence
Padlock_1_MG	/5Phos/CTTTGCAATCAGAAGGTATGATAACAACCCTACGCTCGGAACTAAAATACCCGATAAACGAACGAACGATACCGAACCTAAAACTAATCGAGAAATACCTTGATGGTCAGCAAAA
Padlock_1_HSV	/5Phos/CTACCGGGAGGGGTCGCACCTACGCTCGGATTTCCTACGCTCGGAACTAAAATACCCGATAAACGAACGAACGATACCGAACCTAAAACTAATCCCTACGCTCGGATTTCGGTTGAGTCGGTTTCCTACGCTCGGAGTGTACATGTCCCCGTTTTACGG
Padlock_1_NG	/5Phos/GGAACCCGATATAATCCGCCCTTCGGTTGAGTCGGACTAAAATACCCGATAAACGAACGAACGATACCGAACCTAAAACTAATCCCTACGCTCGGACATATTGTGTTGAAACACCGCCC
Padlock_1_HIV	/5Phos/CCAAATGAGAGAACCAAGGGGAAGCGGTTGAGTCGGACTAAAATACCCGATAAACGAACGAACGATACCGAACCTAAAACTAATCCGGTTGAGTCGGTTTCCTACGCTCGGAGCAGGGCCTATTGCACCAGG
Padlock_1_TP	/5Phos/CTGTCGCACCGTGAGTTCATCCGGTTGAGTCGGTTTCGGTTGAGTCGGACTAAAATACCCGATAAACGAACGAACGATACCGAACCTAAAACTAATCCGGTTGAGTCGGTTTCCTACGCTCGGATTTCGGTTGAGTCGGGAAAAGGGCTGATGGCTCTGAG
Padlock_1_CT	/5Phos/CCTTATGATCGACGGAATTCTGTGGCGGTTGAGTCGGACTAAAATACCCGATAAACGAACGAACGATACCGAACCTAAAACTAATCGTGGGAATCCTGCTGAACCAAG

Synthetic targets	Sequence
Target_1_MG	5’-GTTGTTATCATACCTTCTGATTGCAAAGTTTTGCTGACCATCAAGGTATTTCTC
Target_1_HSV	5’- TGCGACCCCTCCCGGTAGCCGTAAAACGGGGACATGTACACAAAG
Target_1_NG	5’- AAGGGCGGATTATATCGGGTTCCGGGCGGTGTTTCAACACAATATGGCGG
Target_1_HIV	5’- CTTCCCCTTGGTTCTCTCATTTGGCCTGGTGCAATAGGCCCTGCATGC
Target_1_TP	5’- GATGAACTCACGGTGCGACAGCTCAGAGCCATCAGCCCTTTTCAGC
Target_1_CT	5’- CCCACAGAATTCCGTCGATCATAAGGCTTGGTTCAGCAGGATTCCCCACA

Table 3.2. Padlock probe and synthetic target sequences for Digital Ratiometric Fluorescence Coding Assay demonstration

Padlock probes targeting gene sequences in *Microplasma genitalium* (MG, 1420:1473 from M31431), *Herpes Simplex Virus* (HSV, 1426:1473 from K01760), *Neisseria gonorrhoeae* (NG, 362:411 from X52364), *Human immunodeficiency virus* (HIV, 258:305 from KC966998), *Treponema pallidum* (TP, 801:846 from KC966998) and *Chlamydia trachomatis* (CT, 50:99 from JX648604) were designed for the demonstration of our Digital Ratiometric Fluorescence Coding assay.

	Number of binding sites (S)				
Number of colors	4	5	6	…	10
2	7	11	13	…	41
3	22	40	55	…	226
4	51	103	161	…	849

Table 3.3. Approximation of maximum number of fluorescence ratios based on our Ratiometric Fluorescence Coding strategy

The maximum number of fluorescence ratios is calculated given the number of fluorescence colors and fluorescent probe binding sites using our Ratiometric Fluorescence Coding strategy. Adapted with permission from [123]. Copyright 2018 American Chemical Society.

Chapter 4

4. Single-molecule Fluorescence Spectroscopy for Telomere Length Measurement

4.1 Background

Telomeres are specialized DNA tandem repeats and protein structures that cap the chromosomal ends and protect the chromosomes from degradation by serving as sacrificial bases during each cell replication. The rate of attrition of these tandem repeats is around 50 to 200 base pairs with every cell division [124]. When telomeres become critically short, they trigger DNA checkpoint responses mediated by telomere-associated proteins that prevent further cell replication (**Figure 4.1A(i)**) [125-127].

Dysfunction of telomeres can take the form of premature shortening (as in many hereditary telomere syndromes such as Dyskeratosis Congenita, Hoyeraal-Hreidarsson syndrome and pulmonary fibrosis) [125, 126] or lengthening. The latter frequently occurs in cancer, and is preceded by failure to arrest replication in the presence of critically short telomeres [128, 129], and the rescue of ensuing cellular crisis by activation of either telomerase (85-90% of tumors) or proteins associated with alternative lengthening of telomeres (ALT, 10-15% of tumors) [127, 129, 130]. This in turn permits the cells to multiply without constraint.

Telomere length (TL) is also of great interest in the context of the ageing process. However, results from studies using TL to test a host of hypotheses related to the biology of human ageing have often been inconsistent. In using average TL as the only parameter, these studies fail to take into account the heterogeneity of TLs on chromosome arms, first reported by Lansdorp, et al [131]. In fact, it is increasingly recognized that the deleterious effects of telomere dysfunction are mediated by the load of critically short telomeres [132-137]. For example, genetic studies in mice have shown that the shortest telomeres, rather than the average TL, are critical for chromosome stability and cell viability, and are likely a major cause of age-related pathologies [133]. This load can increase due to the gradual shortening of telomeres across all chromosomal arms during normal cellular aging. However, even without significant differences in average TL, the load of critically short telomeres between samples can differ significantly due to variations in the shape of the TL distribution, either between individuals or cell types (**Figure 4.1A(ii)**), or because of bi- or multi-modal distribution that may occur as a result of catastrophic telomere loss [138-140], or species-specific telomere biology (**Figure 4.1A(iii)**) [141]. Therefore, determining the distribution of absolute TLs is of great importance.

Apart from monitoring natural TL changes, the prospect of telomere-lengthening treatment for various medical conditions and prolonging longevity has been raised in recent years. This can be achieved by transfecting cells with telomerase reverse transcriptase (TERT), resulting in significant telomere lengthening [142, 143]. In mouse studies, introducing TERT in adult and old mice resulted in improvements in all aging markers, increased longevity, without any increase cancer incidence [143]. A more recent study also raised the possibility of reducing telomere attrition, or even elongating telomeres in patients with various telomere diseases using androgen therapy [144]. While questions of safety (including the risks of cell immortalization) and efficacy (how and when to introduce the treatment) remain to be answered, these applications argue for the need for a method for determining TL profiles that has a higher throughput.

The strengths and weaknesses of existing TL measurement methods, which have been well documented in literature, are briefly summarized below [136, 145]. Despite their high sensitivity, quantitative polymerase chain reaction (qPCR) methods using primers specific to the telomere repeat sequence yield only average TL for a given sample [57, 58]. Flow-fluorescence *in situ* hybridization (Flow-FISH) fluorescently labels telomeres in blood cells that are then analyzed by flow cytometry and gives the distribution of total TL in each cell [145-147]. Neither method can detect small fractions of critically short telomeres among all chromosomal ends, arguably the most important telomere biomarker. Although terminal restriction fragments (TRF)/Southern blot analysis can in principle yield information of the size distribution of telomeres, it is only semi-quantitative, with relatively poor sensitivity, particularly for shorter telomeres [148, 149]. Furthermore, the size estimates are skewed by the presence of a sub-telomeric sequence of variable size on each fragment, limiting their utility, which we discuss about in the following chapters [149].

A newer PCR-based approach known as single telomere length analysis (STELA) using chromosome-specific primers provides the TL of specific chromosomal arms [145, 150]. However, STELA fails to adequately capture the complete picture of the TL dynamics within cells since primers exist for less than one-fifth of all chromosomal arms [145]. Lastly, quantitative FISH (q-FISH) is able to provide an abundance of information, including chromosomal arm-specific TL. However, the method requires cells that can be induced into metaphase, and thus precludes its application to archival samples such as frozen DNA samples. Furthermore, sample preparation is very time consuming and labor intensive, thus severely limiting its practical applications [145].

Against this technical backdrop we have developed a method dubbed peptide nucleic acid (PNA) Hybridization and Analysis of Single Telomere (PHAST) assay to determine the length of telomeres, hybridized with fluorescent probes, via analysis of single telomere fluorescence measured by a light-sheet fluorescence spectroscopic method [68, 69]. Starting with genomic DNA, the entire sample preparation procedure takes less than 3 hours, including a 2-hour PNA hybridization interval. Telomere molecules are detected one at a time as they transit an observation volume (*OV*), generating single-molecule fluorescence bursts. The properties of these bursts are then analyzed and used to gate the telomere populations, in an approach akin to flow cytometry, to remove spurious bursts and improve the detection limit. By leveraging the unique capabilities of single molecule detection, we have been able to determine the telomere size distribution with DNA from small numbers of cells

4.2 Experimental Details

4.2.1 CICS Instrumentation

The instrumentation for CICS is identical to that previously reported [68, 69]. Briefly, an Ar-ion laser (488 nm, Melles Griot, Carlsbad CA) is passed through an optical fiber and additional shaping optics that expand the laser spots into large illumination discs, which are then focused into a light sheet using a cylindrical lens. The power output as the beam exits the coupler is adjusted to 2.5 mW prior to each experiment. The light is then focused into a microfluidic channel using a 100× oil immersion microscope objective (UPlanFl, Olympus, Center Valley PA). The objective collects the emitted photons from the sample in an epifluorescence configuration, and a dichroic mirror (z488/633rpc, Chroma Technology, Bellows Falls, VT) is used to separate the excitation light and emission fluorescence. A confocal aperture (600 × 150 μm, National Aperture, Salem, NH) is used to spatially filter out out-of-plane light, yielding an effective detection volume of 7 × 2 μm on the microfluidic chip. Finally, dichroic mirrors and band-pass filters (Omega Optical, Brattleboro, VT) are used to spectrally separate the green and red fluorescence and to select the desired wavelengths for detection on the avalanche photodiode (APD SPCM-AQR13, Perkin Elmer, Vaudreuil, Quebec, Canada). Fluorescence data is collected and processed using custom software using the Labview interface.

4.2.2 Telomeric Standards Generation

Telomere length standards were synthesized through a repeated extension process. Initial telomere template (90 bp) was synthesized by Integrated DNA Technologies, Inc. (Coralville, IW), with KpnI and XhoI site at 5’ end and SalI at 3’ end containing 12 telomere repeats (GGTACCTCGAGGG-$(TTAGGG)_{12}$-TCGAC). Two additional forward (5’-TGGTACCTCGAGGGTTAGG-3’) and reverse (5’-GTCGACCCTAACCCTAACC-3’) primers that can anneal to the telomere template at both ends were synthesized and used in a PCR reaction to create a double strand telomere template which was then ligated into pCR4-TOPO TA vector (Life Technologies, Grand Island, NY), and sequence verified by Sanger DNA sequencing (Genwiz, Germantown, MD). The telomere template insert was then released from the TA vector by XhoI and SalI digestion. After separation by electrophoresis on a 1.5% agarose gel and band purified with a DNA gel purification kit from Qiagen, the double-strand template was further cloned into the XhoI site of pUC19 (New England Biolabs, Ipswich, MA (NEB)). To use as a subsequent hosting vector, this obtained plasmid was cut with XhoI. To prepare telomere insert for extension, the cloned insert in the plasmid was released with KpnI and SalI, or by amplification with PCR using the plasmid as template and two primers annealing to pUC19 beside the cloned insert (pUC19-F: 5’-AGTGAATTCGAGCTCGGTAC-3’ and pUC19-R: 5’-CAAGCTTGCATGCCTGCAG-3’) followed by digestion with SalI. Finally, the prepared hosting vector and telomere extension insert were mixed and fused together with a Gibson Assembly cloning kit (NEB) to have the telomere insert in the hosting vector extended. This process was repeated several times in order to obtain longer telomere inserts. To check the length, telomere inserts were released with XhoI and SalI digestion and run on a 1.5% agarose gel.

4.2.3 DNA Digestion by Restriction Enzyme

To digest the DNA samples, we incubate 20 µL of the DNA (typically between 1-5 µg total DNA) with 4.39 µL of a mixture of RsaI and HinfI restriction enzymes (Roche Applied Sciences, 1000 units) and Buffer A (Roche) (1:1:2 volume ratio) at 37 °C for 2 hours.

4.2.4 Single-tube Purification and Hybridization

To minimize the DNA fragmentation during telomere preparation, we chose to apply phenol/chloroform extraction, which is considered as the gold standard to produce high-purity and high-molecular weight DNA with minimal shearing. Genomic DNA is extracted from cell lines using phenol/chloroform extraction and added to a 150 µL proprietary PNA hybridization buffer containing 500 pM of telomere-specific biotinylated PNA probes (pcTel-Bio, PNA Bio Inc.). After a 10 minutes incubation at 85 °C, the mixture is left at room temperature for 30 minutes. 20 µL of streptavidin-coated magnetic Dynabeads C1 (Life Technologies) is then washed in 1X Binding Washing (BW) buffer as described by manufacturer, resuspended in 150 µL 2X BW buffer and incubated with the DNA/PNA mixture for 30 minutes on a rotating sample holder. The pulled-down DNA is then washed with 1X BW buffer to isolate the telomere-containing DNA.

4.2.5 Fluorescent PNA Hybridization

DNA samples in solution are mixed directly with 500 nM Alexa 488 labeled PNA (PNA-AF488, PNA Bio Inc.) in PNA hybridization buffer, incubated at 85 °C for 10 minutes and left at room temperature in the dark for 2 hours to complete the labeling. For DNA samples bound to magnetic beads, the telomeric DNA is released from the pcTel-Bio/Dynabeads upon heating and placed on a magnetic holder prior to incubation at room temperature to isolate the telomeres from the beads. The large excess of PNA-AF488 ensures minimal re-hybridization onto the beads. Finally, the samples are purified with S400-HR microspin columns (GE Healthcare) to remove the excess PNA-AF488.

4.2.6 Telomere Sizing on CICS

Microfluidic devices are fabricated as described previously [68, 69, 151]. The purified samples are flowed through a 5 µm x 0.5 µm microfluidic channel (w x h) using a pressure controller (Alicat Scientific) and detected on CICS. Emitted photons are counted at 0.1 ms intervals.

To ensure representative data, the pressure is adjusted to ensure that the average transit time of the samples through the CICS detection volume is at least 3 ms. Gating to exclude spurious peaks is performed as described in **Chapter 4.6**. A standard curve is constructed using linearized cloned telomere standards (0.1, 0.2, 0.4, 0.9, 1.2 and 2.4 kb TL), relating the height of the fluorescence peak to the length of the DNA. We have found that the fluorescence peak height varied linearly with TL and is invariant between experiments. Therefore the same standard curve is used for size estimation. For TLs exceeding the range of the standards, a linear extrapolation is used to estimate their sizes as discussed in the **Chapter 4.8**. In addition, 3 µg of HeLa genomic DNA (NEB) was processed using our PHAST assay as described above and the resultant telomere length was compared with that measured by TRF (**Chapter 4.9**).

4.2.7. Determination of Pull-down Efficiency

To determine the efficiency of the pull-down assay, a mixture of telomeric standards (0.2, 0.4, 1.2 and 2.4 kb) at high (1.60, 1.64, 1.84, 2.16 ng respectively) and low concentrations (40, 41, 46, 54 pg respectively) was mixed with 1µg of heat-fragmented genomic DNA (~ 100-1000 bp) prior to sizing with PHAST assay (200 µL final volume). As controls, telomeric standards was mixed at the same proportions and hybridized with PNA-AF488, without going through the pull-down process. The pull-down efficiency of each species is then determined by comparing the number of molecules detected with and without pull-down.

4.2.8 Comparison of PHAST with q-FISH

WI38, U2OS and R83 cells were purchased from ATCC (Manassas, VA) and cultured in the medium as recommended by the vendor. Human primary fibroblasts were the gift of Drs. Xuefeng Liu and Richard Schlegel in the Molecular Oncology Programs, Lombardi Comprehensive Cancer Center, Georgetown University. The cells were cultured in DMEM plus 10% FBS, and harvested at metaphase, followed by chromosome preparation according to the method described previously [152]. The same batch of cells in separate flasks were harvested with 0.1% Trypsin/EDTA (Invitrogen, Rockville, MD), and genomic DNA from cell pellet were purified with DNA purification kit (Qiagen, Germantown, MD). The quality and concentration of the DNA was measured with Nanodrop.

Chromosome arm-specific telomere length was measured by telomere quantitative fluorescence in situ hybridization (q-FISH) as previously described [153]. Briefly, chromosome preparations were dropped onto clean microscopic slides and hybridized with 15 μL of hybridization mixture consisting of 0.3 μg/ml Cy3-labeled telomere-specific PNA probe (PNA Bio Inc.), 1 μL of cocktails of FITC-labeled centromeric PNA probes specific for chromosomes 2, 4, 8, 9, 13, 15, 18, 20 and 21, and 20 μg/mL of Cy3-labeled centromeric PNA probes specific for chromosome X (Biomarkers, Rockville, MD), in 50% formamide, 10 mM Tris-HCl, pH 7.5, and 5% blocking reagent. Slides were denatured and then hybridized at 30°C for 3 hours. After hybridization, the slides were sequentially washed 10 minutes each at 42°C: once in 1 x SSC, once in 0.5 x SSC, and once in 0.1 x SSC. The slides were then mounted in anti-fade mounting medium containing 300 ng/ml DAPI.

After q-FISH, cells were analyzed using an epifluorescence microscope equipped with a charge-coupled device (CCD) camera. Metaphase cells were captured with exposure times of 0.15, 0.25 and 0.05 second for Cy3, FITC and DAPI signals, respectively. Digitized metaphase images were analyzed using the Isis software (MetaSystems Inc. Boston, MA), which permits the measurement of 92 telomere signals simultaneously after karyotyping. Telomere fluorescence intensity units (TFU) were recorded as an indirect measurement of telomere length. For each sample, at least 12 metaphase cells were analyzed. Each pair of homologous telomeres was recorded separately as homologous short (S) and homologous long (L).

PHAST assay was performed on three cell lines, namely WI-38, R83 and U2OS. A total of around 5 µg genomic DNA is processed as previously described. q-FISH results were obtained from our collaborators at Georgetown University. Since q-FISH does not provide absolute telomere length, the TFU are scaled to match the mode of the TFU histogram to the telomere length histogram calculated using PHAST.

Similarly, PHAST assay was performed on three clinical samples (two from cancer patients, one from a control subject). Genomic DNA was isolated from three clinical blood samples (BA01628, BA01650, CB3319) obtained from our collaborators at Georgetown University. For each clinical sample, a total of around 500 ng genomic DNA was processed as described previously. Q-FISH results were also obtained from our collaborators who offered the clinical samples for comparison.

4.2.9 Telomeric Sequence Generation via Rolling Circle Amplification

It is difficult to obtain telomere standards longer than 2.4 kb because the cloning vector is unable to accept a much larger cassette. As a result, we employed padlock probe chemistry [106] and Rolling Circle Amplification [107] to produce longer telomeric sequences. We design our padlock probe to contain two non-telomeric regions at 5′ and 3′ ends, and an internal part consisting of 9 repeats of CCCTAA sequences. In the presence of oligo sequence complementary to the non-telomeric regions, padlock probe is ligated to a closed circle for subsequent amplification and result in single strand DNA over 20kb, a substantial proportion of which are telomeric sequences. The sequence of our padlock probe is 5'/Phos/-GGAACCCGATATAATCCGCCCTTCCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAAATATTGTGTTGAAACACCGCCC. The sequence of RCA primer is 5'-GGGTTAGGGTTAGGGTTA and the sequence of complementary non-telomeric oligo is 5'-AAGGGCGGATTATATCGGGTTCCGGGCGGTGTTTCAACACAATATGGCGG. All DNA oligonucleotides were purchased from

Integrated DNA Technologies, Inc. (Coralville, IA). Reagents for ligation and RCA reactions, including 9°N™ DNA Ligase, 10× 9°N™ DNA Ligase Reaction Buffer, Phage ø29 DNA Polymerase, Exonuclease I (*E. coli*), Exonuclease III (*E. coli*) and 10× Isothermal Reaction Buffer were purchased from New England BioLabs, Inc. (Ipswich, MA).

Padlock probe hybridization and ligation was achieved in a 25-µL reaction mixture contained 150 nM of padlock probe, 400 nM of complementary non-telomeric oligo, 0.4 Units/µL 9°N™ DNA Ligase, and 1× 9N DNA Ligase Reaction Buffer (10 mM Tris-HCl, 600 µM ATP, 2.5 mM $MgCl_2$, 2.5 mM Dithiothreitol, 0.1% Triton X-100, pH 7.5 at 25°C). Padlock probe hybridization and ligation was performed at 60 °C for 1 hour. Then 25 µL of the ligation product was treated with 10 units of Exonuclease I and 50 units of Exonuclease III at 37 °C for 1 hour to remove extra padlock probes that were not circularized. After digestion, 5 µL of the enzyme-treated ligation product was mixed with 200 nM RCA primer, 400 µM dNTPs, 10 U/µL Phage ø29 DNA polymerase, and 1× Isothermal Reaction Buffer (50 mM Tris•HCl (pH 7.5), 10 mM $MgCl_2$, 200 µg/mL acetylated BSA) to a final volume of 25 µL. Of note, due to its high strand-displacing activity, Phage ø29 DNA polymerase was added last to the reaction mixture, which was then immediately incubated at 31 °C for 1 hour to perform RCA, followed by 70 °C for 10 minutes to inactivate the polymerase and stop the reaction.

4.2.10 Application to Huntington CAG Tandem Repeats

To apply our method for Huntington CAG tandem repeat, a 12nt Alexa 488 labeled PNA (5'-Alexa488-CTGCTGCTGCTG, PNA Bio Inc.) was used to label the CAG trinucleotide repeat oligonucleotide. The CAG repeat oligonucleotide contains 66 CAG trinucleotide repeats ($(CAG)_{66}$, Integrated DNA Technologies, Inc.), which allows the hybridization of a maximum of 16 PNA probes. The protocol for fluorescent PNA hybridization is similar to that in our PHAST method described above. 500 pM CAG repeat oligonucleotides were mixed directly with 500 nM PNA in

PNA hybridization buffer, incubated at 85 °C for 5 minutes and left at room temperature in the dark for 2 hours to complete the labeling. Then the purified sample was flowed through the microfluidic channel and detected on CICS.

4.3 Overview of PHAST

The PHAST protocol begins with total genomic DNA extracted from cells using standard phenol/chloroform extraction, taking care to avoid shearing the DNA molecules (**Figure 4.1B**). Telomeric DNA is pulled down using biotinylated PNA probes that hybridize to the telomere sequence, and streptavidin-coated magnetic beads. After the non-telomeric DNA has been discarded, the sample is heated to release the telomeres from the beads in the presence of excess Alexa488-labeled PNA (PNA-AF488). The magnetic beads are then discarded and the labeled telomeres are ready to be analyzed.

The sample is introduced into a microfluidic device on a custom-built microfluidic single molecule detection (SMD) instrument as described previously (**Figure 4.1C**) [68, 69, 151]. Briefly, the microchannel spatially confines each passing PNA-AF488 labeled telomere into the path of a laser beam, in a manner analogous to the sheath/core flow configuration in flow cytometry. Typical SMD instruments utilize a diffraction-limited laser spot with a Gaussian profile to excite the fluorescently labeled targets [154]. As a result, the brightness of each single molecule fluorescence burst depends on both the number of dyes on the molecule, as well as the excitation intensity through which it traverses. The inability to decouple these two contributing factors yields highly variable fluorescence signal from even a uniform sample, and limits SMD's ability to characterize individual molecules. To address these limitations, PHAST uses a cylindrical illumination confocal spectroscopy (CICS) configuration, wherein the excitation illumination is modified using beam-shaping optics to give an *OV* in the shape of a uniform light sheet that extends across the entire microchannel cross-section (**Figure 4.1C**) [68, 69, 151]. Not only does this allow quantification of the number of PNA-AF488 probes on the telomeres, but it also ensures that every single telomere molecule passing through the microchannel is accounted for, thereby eliminating sampling bias, and in principle permitting the use of very few cells for quantification.

4.4 Single-molecule Fluorescence Peak Parameters Modeling

Detection of the fluorescently labeled telomere is performed in a microfluidic channel as described previously [68]. Briefly, a solution containing the labeled telomeres is driven through the microchannel using a pressure source at various settings. An expanded laser (488 nm Argon ion laser) confocal spot spans the entire cross-section of the microchannel, and as each molecule passes through this region, it is excited and the emitted photons are collected by an avalanche photodiode (APD). The fluorescence signals collected from the molecules at 0.1 ms intervals are processed using custom Labview software. ND filters are applied when more than 10% of the peak heights exceed 1,000 photon counts per 0.1 ms interval in order to prevent detector saturation. We define a peak (or burst) as having been detected when the fluorescence intensity exceeds a pre-determined threshold, t (= 100 and 40 photon counts when no ND and ND 0.5 is applied to the laser source, respectively) (**Figure 4.2A**). To minimize measurement artifacts arising from fluorescence fluctuation around the threshold, the beginning and end of each peak are signified when the fluorescence drops below the baseline value, b (= 40 and 15 for no ND and ND 0.5). The having different values for t and b ensure that signal fluctuating around the threshold will not register as multiple peaks.

For each fluorescence peak, the time interval between the points at which the signal crosses a lower baseline level along the rising and falling edges is defined as the *Width* of the burst. The maximum intensity and total fluorescence (less the baseline) of each burst are the *Height* and *Size*, respectively. This thresholding procedure ensures that a fluorescent molecule, which barely exceeds the threshold, will not register as multiple events, were its signal to fluctuate around the threshold. While early studies involving CICS used the *Size* exclusively for analysis, recent development has extended its capabilities to other parameters such as the shape of the fluorescence bursts, in the form of packing density of DNA in buffers of different ionic strength [155], and the analysis of fluorescence distribution of entire sample populations [151].

Here the peak parameters have been modeled for identifying true signal bursts from background fluctuation. The laser excitation and detection volume, referred to as the observation volume (OV) of CICS represents the space in which fluorescent molecules will be excited and detected, and is uniform across the channel (x-axis) and along the optical axis (z-axis), and Gaussian along the flow axis (y-axis) [68]. As each molecule traverses this Gaussian OV along the flow axis, the number of photons or fluorescence intensity, I, detected can be modeled as a convolution of the OV with the distribution of fluorophores along the flow axis (Φ (y, L)), of the general form

$$I(y,L)=\Phi(y,L)\otimes OV(y) \tag{1}$$

Single fluorophores and short DNA sequences hybridized to fluorescent PNA probes can be modeled as point emitters, i.e.

$$\Phi(y,L)\propto \delta y \tag{2}$$

$$\Phi(y,L)=\kappa\bullet L\bullet\delta y \tag{3}$$

where κ and L represent a constant of proportionality and the length of the DNA, respectively. This assumes that the length of the DNA is directly proportional to the number of bound probes, and hence fluorescence. The resulting intensity function, $I\ (y,\ L)$, thus assumes the shape of the OV, *i.e.* Gaussian, where

$$OV(y)=e^{\frac{-y^2}{2c^2}} \tag{4}$$

$$I(y, L) = \kappa \cdot L \cdot e^{\frac{-y^2}{2c^2}} \tag{5}$$

where c is the parameter describing the size of the OV, and has been determined to be 0.125 microns previously [68].

On the other hand, longer DNA molecules can be modeled as approximately spherical (if unconfined within the microchannel) or ellipsoidal (if confined) finite-sized random coils with the fluorescence homogeneously distributed in the corresponding volumes. In the first case, the radius of gyration, R_g, is estimated as

$$R_g \propto L^{\gamma} \tag{6}$$

$$R_g = \beta L^{\gamma} \tag{7}$$

The β value (0.0412) is derived from reported hydrodynamic radii for linear DNA [118]. The γ value (0.6) is based on the well-established Flory approximation, assuming a free polymer chain in a good solvent [117, 156]. For a channel with a depth, d, of 0.5 micron tall channel, the onset of weak confinement on the DNA free coil is around 8-9 kb, corresponding to $R_g \approx 0.3d$, or 150 nm [117]. Even 50-60 kb DNA molecules (representing the typical maximum DNA length that can be extracted from cells) experience only moderate confinement ($R_g < d$) [117, 157]. As such, we will restrict the analysis to spherical approximations of the DNA molecules.

$$V = \frac{4}{3}\pi R_g^3 = \frac{4}{3}\pi \beta^3 L^{3\gamma} \tag{8}$$

where V is the total volume of DNA molecules. Since the total number of fluorophores is proportional to the length of DNA, we can define total fluorescence of each DNA molecule, F, and its spatial density, or fluorescence density, f_v,

$$F(L) = k \bullet L \tag{9}$$

$$f_v(L) = \frac{F(L)}{V} = \frac{k \bullet L}{V} = \frac{3kL^{1-3\gamma}}{4\pi\beta^3} \tag{10}$$

In this model, *L* represents the total length of the fluorescent segment, while *k* is a scaling parameter to account for both the intrinsic brightness of the fluorophore, as well as the intensity of the excitation source.

The incremental volume, *dV*, along the flow axis is parabolic for both unconfined and confined DNA random coils, with different coefficients. The distribution of mass along the flow axis can thus be represented by *ρ(y,L)*, where

$$\rho(y,L) = \begin{cases} \pi(R_g^2 - y^2) = \pi((\beta L^\gamma)^2 - y^2), & \mathit{for} \quad -\beta L^\gamma < y < \beta L^\gamma; \\ 0 & \mathit{otherwise} \end{cases} \tag{11}$$

$$dV = \rho(y,L) \bullet \delta y \tag{12}$$

Assuming that the fluorophores are homogeneously distributed in this volume, the fluorophore distribution, Φ(*y*,*L*), is the product of this parabola (*dV*) and the fluorophore density (*f*_v*(L)*).

$$\Phi(y,L) = \rho(y,L) \bullet f_v(L) \bullet \delta y = \frac{3kL^{1-3\gamma}}{4\beta}\left(1 - \left(\frac{y}{\beta L^\gamma}\right)^2\right) \bullet \delta y, \quad -\beta L^\gamma < y < \beta L^\gamma \tag{13}$$

The peak profile, *I (y, L)*, will then be simply the convolution between the fluorophore distribution, Φ(*y, L*), and the Gaussian OV profile OV(*y*) (**Figure 4.3**).

As the radius of the DNA increases, the peak profile transitions from Gaussian (when $R_g << c$) to parabolic (when $R_g >> c$), as the OV and *Φ* functions dominate the peak profile function, respectively. It is worth noting here that while the horizontal axis in **Figure 4.3** is in terms of position (*i.e.* y-value), it can be converted into a time axis if the velocity is known since the APD samples the fluorescence at fixed time intervals.

The DNA diameter corresponding to the largest telomeres (~ 0.3 microns for 10 kb molecules) is comparable to the size of the Gaussian OV, which has a $1/e^2$ diameter of around 0.25 micron along the flow axis [68]. Consequently, the peak profile for long telomeres is intermediate between Gaussian and parabolic, for which no simple analytical solution exists. Nevertheless, numerical approximations of the peak profile for different DNA lengths and the corresponding parameters can be calculated using mathematical software (Wolfram Mathematica 9).

S-values are obtained by integrating the peak shape over the time interval, *W*, which is determined using the same threshold that is used for the experimental data. To obtain the peak height or *H* from this model, it is noted that the maximum signal occurs when the Φ and *OV* functions are centered, *i.e.* when the DNA lies at the center of the OV, such that $I\ (0, L)$. To recapitulate the peak shape using the model, only two variables have to be chosen empirically using our experimental data. The first is the *k*-value, which is a single factor to account for excitation intensity, detector efficiency, DNA labeling density, and dye quantum yield. The second is a scaling factor applied to both *S* and *W* to account for velocity.

Comparing the peak heights and widths from the model with experimental results obtained from Hind III digested λ DNA labeled with TOTO-3 indicates a good fit between the two (**Figure 4.4**). The λ DNA digest system was chosen because the sizes of the fragments are well defined, and the longest fragment is larger than 20 kb, as opposed to only 2.4 kb for our telomere standards. This allows us to examine the accuracy of the model when deviating from the point emitter idealization. Furthermore, since any difference between the Hind III digest and Alexa488-labeled telomere systems can be accounted for by adjusting the *k* (incident intensity, dye quantum yield, and labeling density) and *c* (width of the Gaussian OV) factors, there are no explicit reasons to expect significant differences between this and the telomere system. Thus, the model should be able to account for both these systems with minimal change. Derived essentially from first principles, the extensive agreement between the model and experimental data suggests that the model does indeed describe the physical system, and can be used to inform our gating strategy (details in **Chapter 4.6**), and even predict the peak parameters.

4.5 PHAST One-Tube Purification Assay

For a typical sample, total genomic DNA contains a large amount of non-telomeric sequences. In fact, telomeres constitute around just 0.01-0.02% of the total DNA mass. As a result, even though only a small proportion of genomic DNA will hybridize non-specifically to the Alexa 488 labeled PNA (PNA-AF488) probes, the sheer abundance and different sizes of these molecules makes them difficult to remove by size selective methods (dialysis, microspin columns, *etc*). Because each DNA molecule is expected to bind very few PNA probes in this manner, the presence of these non-specifically labeled molecules manifests as a generally higher background, which increases the chance of erroneous detection. Furthermore, since we define the start and end of peaks as the points at which the photon count crosses the baseline level (**Figure 4.2B**), if the background were to approach or exceed the baseline, it will cause errors in the measurement of all peak parameters, since the demarcation between one peak and the subsequent one(s) may not be possible to ascertain.

To address this, we developed a one-tube purification assay to allow removal of virtually all of the non-telomeric DNA molecules (**Figure 4.5**). Briefly, telomere-containing DNA molecules are hybridized with biotinylated telomere-specific PNA capture probes (pcTel-Bio) without any prior digestion, then immobilized on streptavidin-coated magnetic beads, allowing the non-telomeric DNA fragments-containing supernatant to be removed (**Figure 4.5, Steps (I)-(III),**

Panels A and B). After rinsing, the captured telomeres are released into a solution of PNA-AF488 probes at an elevated temperature. By using 100 times as much PNA-AF488 probes as biotinylated ones, we ensure that the fluorescent probes will compete favorably against any biotinylated capture probes that remains in solution (**Figure 4.5, Steps (IV)-(V), Panels C and D**). This whole process is performed in a single tube, and thus further development such as automation and implementation on a microfluidic platform is uncomplicated.

To demonstrate the effectiveness of the pull-down protocol, 45 pg of 2.4 kb telomere standard in 1 µg of genomic DNA is processed. The gel image (**Figure 4.6A**) shows the intense fluorescence due to the genomic DNA, which completely obscures the telomere band. After the pull-down procedure, the telomeric band can be clearly seen (white dashed box). Non-specific pull-down of long genomic DNA fragments invariably occurs. However, due to the specificity of the biotinylated PNA probes, these fragments are not in sufficient concentrations to show up in the CICS trace (**Figure 4.6B, C**). The pull-down step greatly reduces the background fluorescence, and thus improves the performance of PHAST.

To ensure that the pull-down purification process does not bias the TL distribution, telomere standards at known ratios (total telomere mass of 180 pg) are spiked into heat-fragmented genomic DNA (~100-1000 bp, 1 µg), and purified using our procedure. At this relative telomere concentration, which is comparable to that expected from mammalian genomic DNA, telomeres were recovered in a size-independent manner (~ 50% for all tested TLs, **Figure 4.6D, blue bars**).

However, when much higher concentrations of telomere standards were used (16 ng in 1 µg), we found that the relative pull-down efficiency of the DNA increases with the length of the telomere (**Figure 4.6D, red bars**). This observation can be explained by noting that, in general, we expect the number of biotinylated probes bound to the telomeres to be proportional to the TL. Furthermore, assuming that streptavidin/biotin binding is non-cooperative (a reasonable assumption since it is unlikely for adjacent biotinylated PNAs to bind to the same streptavidin molecule), and since a single successful tether is sufficient to capture a DNA, the association rate of the telomeres will scale with the number of bound biotinylated PNA [158]. Because streptavidin/biotin conjugation can be considered to be an essentially irreversible process, the relative pull-down efficiency for telomeres of different lengths will depend only on the association rate, though partially mitigated by the differences in diffusivity and steric hindrance. If the number of streptavidin is insufficient to pull down all the telomeres, the faster binding rate of longer telomeres with more biotin hybridized to them ensure that they will occupy a larger portion of the available surface.

4.6 Gating Procedure for Background Removal

detect the shortest telomeres, the threshold is lowered to improve the detection limit. This increases the thresholding artifacts, which occur when the background noise drifts above the threshold, tend to hover around the baseline, such that

$$Size_{\text{artifact}} \sim (Width_{\text{artifact}}/\text{interval}) \times \text{baseline} \tag{14}$$

As a result, the $\log_{10}$(width)/$\log_{10}$(size) slope for the artifacts is approximately 1. On the other hand, based on the model described previously, it is determined that the peak width increases fairly slowly with DNA length and hence peak size ($\log_{10}$(width)/$\log_{10}$(size) slope ~ 0.2 for the conditions tested).

These characteristics have been exploited to formulate the gating strategy for the peaks. By clustering the peaks, two subpopulations can be clearly identified (**Figure 4.2C**). As can be seen in **Figure 4.7**, a simple threshold in any of the parameters (burst *Size*, *Width* or *Height*) will fail to adequately separate the actual and spurious peaks. On the other hand, a two-dimensional (width/size) gate can separate them (**Figure 4.7A**, red dashed box). Selecting the cluster with the smaller slope allows us to include many more peaks than merely using a higher threshold, while also excluding the spurious ones. All the cell and tissue samples were gated in this manner and compared to the q-FISH results.

4.7 Single-molecule Fluorescence Burst Analysis of Telomeres

In PHAST, events of single-molecule fluorescence bursts of telomeres are identified from the raw fluorescence signals using a predetermined threshold (**Figure 4.2A**). The parameters (*i.e. Size, Width* and *Height*) are employed, together with a theoretical framework based on the model of peak parameters described above to formulate a strategy for isolating true signal bursts from background fluctuation. Depending on the background noise level, some spurious bursts will invariably be acquired due to signal fluctuation (**Figure 4.2B**). Such peaks can be efficiently removed by selecting a sufficiently high threshold at the expense of assay sensitivity. However, since the abundance of the shortest telomeres are of greatest interest, our chosen threshold, which corresponds to TLs of as little as 65 bp, is low enough to pick up a significant number of spurious peaks, particularly in the presence of relatively large quantities of residual non-telomeric DNA (as discussed in **chapter 4.5**). These real and spurious peaks form two distinct clusters in a *Width*/*Size* plot (**Figure 4.2C**). This is because the background drifts that cause the spurious peaks typically hover around the baseline, with a nominal burst *Size* that is proportional to *Width*, *i.e.* the spurious cluster will have a slope of ~ 1 on a log-log plot of the two parameters. On the other hand, the physical constraints of the *OV* size, DNA length, and flow velocity place limits on the possible *Width* values of peaks associated with the detection of actual fluorescent molecules, and as such the corresponding cluster will have a gentler slope on the same plot. This is corroborated by the theoretical estimates (slope ~ 0.2, as described in chapter **4.4**). Based on these distinguishing features, a gate is constructed that can isolate the spurious cluster (**Figure 4.2C, D**, **chapter 4.6**).

One limitation of previous implementations of CICS is that, since the total number of photons in a burst is proportional to the residence time in the laser excitation volume, comparison between different samples can only be made if the flow velocity is accounted for. Furthermore, because the pressure-velocity relationship depends on a number of factors, including microchannel geometry variation, surface properties, *etc*. the actual velocity in each experiment cannot be determined easily. By analyzing the parameters of telomere standards of different lengths (0.2, 0.9 and 2.4 kb) run through the PHAST instrument at different pressures, the effects of velocity on PHAST parameters are apparent (**Figure 4.2E, F**). Positions of the clusters in the *Width/Size* plot can be decoupled into two components, along the Flow Velocity Axis and Telomere Length axis. Since both *Width* and *Size* scale linearly with flow velocity, the Flow Velocity Axis has a slope of ~1. This allows the same gate defined previously for removing spurious peaks to be used regardless of flow velocity. In fact, we found that the flow velocity had minimal effect on the burst *Height*, with only a 10% change when the flow velocity is increased 8-fold (**Figure 4.2F**). This translates to a smaller coefficient of variation (CV) for *Height* (10%) compared with *Size* (25%) in the length estimation of a 2.4 kb standard. Therefore *Height* is chosen as the parameter for telomere length estimation. Given the parallels between this CICS-based fluorescence spectroscopic approach and flow cytometry, we have named it fluorescence flow moriometry (Gr. *mório*, molecule or small particle, + *metron*, to measure), of which PHAST is one of many applications.

4.8 Measurement of Telomere Standards by PHAST

Previous implementations of CICS utilized peak burst size for quantification. Since peak size changes with flow velocity (slower moving molecules will reside in the OV longer, resulting in larger number of total detected photons), a calibration step is necessary to account for different flow speeds. Although the flow velocity varies linearly with the driving pressure, the exact linear relationship (the slope, *x*- and *y*-intercepts) changes from chip-to-chip, depending primarily on the hydrophobicity of the PDMS microchannel (which increases with time after oxygen plasma bonding with glass), and to a lesser extent the variations in the dimensions of the microchannels.

Since we can only control the driving pressure directly, there is no way to be certain of the flow velocity without utilizing flow tracers such as fluorescent microspheres [68, 69, 151].

Flow tracers typically require a separate fluorescence channel (that is, an additional laser excitation source and emission detector) that is coincident with the sample channel. This is non-trivial since the alignment of the optical components has to be extremely close. Since light of different wavelengths focus at slightly different positions along the optical axis, it is necessary to record the two spectral channels sequentially, with an additional focus adjustment step in between. Doing away with the flow tracers greatly simplifies the demands on the instrumentation. For the existing CICS instrument, a tracer-free CICS also frees up an additional spectral channel to permit simultaneous multi-color detection.

In our model, changes in flow velocity merely broaden peaks proportionately, resulting in a simple scaling of the peak sizes and widths without any change to the shape of peaks. This was borne out by experimental data where the slope of $\log_{10}$(width)/$\log_{10}$(size) curve for each telomere standard hybridized with Alexa 488-labeled PNA at different driving pressures has a slope of around 1 (**Figure 4.2E**). Furthermore, since the $\log_{10}$(size)/$\log_{10}$(width) curve for each velocity has the same slope (~0.2) and thus never intersects with another, a characteristic curve corresponding to each flow velocity can be obtained. Consequently, each $\log_{10}$(size)-$\log_{10}$(width) co-ordinate corresponds to a unique velocity and telomere length (TL), *i.e.* for a given set of experimental conditions, we expect all the real peaks to lie parallel to the TL axis (**Figure 4.2E**).

Peak size (*i.e.* total photons detected during the transit of a fluorescent molecule through the detection volume) is, in principle, the best measure of the actual TL, once the flow velocity is accounted for. This is because each fluorophore has the same path integral in CICS, and hence the total number of detected photons corresponds to the total fluorophore content, and by extension TL, regardless of the spatial distribution of the fluorophores (*i.e.* peak shape). However, fluctuations in the flow velocity are inevitable, which introduce additional variation to the peak size. This variation can be observed in **Figure 4.2E**, where each cluster of peaks is elliptical, with its long axis approximately parallel to the Flow Velocity axis. A transformation can be performed to convert the flow velocity and TL axes into an orthogonal pair for simple telomere sizing. Alternatively, given the consistent ellipsoidal molecular shape for all the expected DNA molecules (as discussed in the previous chapters), peak height can also be utilized for telomere sizing.

The model predicts that in general, the peak height is a good proxy for each TL that is independent of flow velocity, except when the velocity is high enough to render the sampling rate inadequate, or if the DNA molecule is much larger than the *OV*. The former can be seen in minimal shifts in peak heights (~10%) with an 8-fold change in flow velocity when the driving pressure was increased from 0.25 to 2 psi (**Figure 4.2F**). Therefore, by simply flowing the samples through the microchannels at adequately low pressures (~ 0.5 psi), it is possible to obtain good TL estimates without correcting for the exact velocity.

Because most of the DNA molecules are small compared with the Gaussian OV and hence well described by the scaling parameter κ (**Chapter 4.4**), peak height also varies linearly with TL for relatively short telomeres (< 10 kb). On the other hand, very long DNA molecules deviate from the point-emitter idealization. In these cases, the fluorophores are much more spatially distributed, and only part of the DNA lies in the detection volume at any point in time, and thus results in a decrease in the maximum instantaneous fluorescence (*i.e.* the peak height). While the telomere standards are too short for us to observe this drop-off (the longest telomeric segment is only 2.4 kb long), both the initial linear relationship and the subsequent drop-off were observed experimentally for Hind III digested lambda DNA (**Figure 4.4**). Although this nonlinearity presents some uncertainty in obtaining TL values from peak height by extrapolation, for TLs up to 10 kb in length, a linear approximation is deemed sufficient. Even for telomeres up to 30 kb in length, the linear approximation underestimates TL by only 15-20%. Therefore, the peak height is considered adequate as a general sizing approach. Based on our experiments, the conversion from peak height, H, to TL can then be performed by a simple linear transformation, where

$$TL\ (\text{kb}) = (H - 50)/780 \tag{15}$$

for Alexa488 labeled probes.

Based on the values derived from telomere standards of known lengths (0.1 to 2.4 kb), burst *Height* is related to telomere length through a simple linear relationship (**Figure 4.8A**). The TL estimates has a CV of around 10-25%, with greater uncertainty for shorter telomeres (**Figure 4.8A, B**), as well as very good reproducibility for three independent sets of standards tested at 6-month intervals (CV < 11%, **Figure 4.8A**, **Table 4.1**). Although we expect the linear relationship to break down for much longer telomeres, it remains a valid approximation for up to 10-15 kb telomeres (the average TL of neonates) [131, 147], and underestimates the actual TL by only 15-20% for 30 kb telomeres, which is far longer than telomeres normally found in humans. To ensure that the gating procedure does not skew the TL estimates, a 2.4 kb telomere sample was tested with and without genomic DNA background. The resulting TL distribution was found to be unaffected by the gating procedure (**Figure 4.8C**).

Considering the fact that the size of our telomere standards is limited to 2.4 kb because the cloning vector is unable to accept a larger cassette, we used padlock probe chemistry in conjunction with rolling circle amplification to generate long DNA molecules containing telomeric sequences to simulate long telomeres (**Figure 4.9A**). Like real telomeres, each sample prepared will contain DNA molecules of different lengths. By subjecting the sample to the pull-down process and comparing the telomere distribution determined by PHAST before and after the pull-down, we determined that our process did not bias the telomere size distribution, as evidenced by the linearity of the Q-Q plot (**Figure 4.9B**). Furthermore, it is worth noting that some of the telomeres detected here were longer than 10 kb. This is comparable to the telomere lengths found in cells, and the absence of bias confirms the veracity of our length estimates.

4.9 Comparison between PHAST and TRF

Although it has long been known that critically short telomeres can have an outsized impact on cell phenotype, what constitutes this critical threshold has, frustratingly, been poorly defined. This is primarily due to the absence of tools that can reliably size short telomeres. We have found only one study that determined, using STELA and sequencing, that XpYp/17p fusion occurs below a minimal threshold of 13 hexanucleotide repeats, which is needed for telomeres to function normally [137]. However, the method is not widely used, possibly due to the laborious and specific nature of the methodology. Nevertheless, the study confirms that a physical limit on functional TL exists, below which the protective functions of telomeres are lost, perhaps due to insufficient binding of telomeric repeat binding factors (e.g. *TRF1*, *TRF2*) [137]. So far, the only conventional measurement method that can routinely provide estimates for absolute TL is TRF.

TRF followed by Southern blotting is one of the most commonly used methods for telomere sizing. It is an elegant method for measuring the size of telomeres that takes advantage of the persistence of the canonical hexanucleotide (TTAGGG) repeats in the presence of restriction enzymatic activity [148]. It is one of very few methods that can yield any actual absolute telomeric size information, and is particularly useful when comparing the changes to the TL in a cell population over time. However, since the length of the final telomere-containing molecule depends on the proximity of the nearest restriction site to the telomere sequence, the length of this non-canonical sub-telomeric region varies from chromosome to chromosome (**Figure 4.10A, B**). Consequently, a strong signal on a particular position on the gel may be due to a large number of short telomeres with long non-telomeric regions, or a small number of long telomeres with short non-telomeric regions. TRF/Southern blotting is unable to distinguish between these two cases. On the other hand, telomere sizing by PHAST utilizes only the fluorescence intensity, which varies proportionately to the TL. Furthermore, since molecules are detected one at a time, PHAST can decouple the abundance and length of the telomere-containing fragments, thus yielding the actual TL distribution.

To illustrate this, 2.4 kb telomeric sequence ligated into a plasmid vector was digested with *Rsa*I/*Hinf*I restriction enzyme mix, commonly used in TRF protocols. Gel electrophoresis shows the large shift in electrophoretic mobility that results from the loss of the non-telomeric region after digestion (**Figure 4.10C**). When tested with our method, the samples before and after digestion are virtually indistinguishable (**Figure 4.10D**). Since extracted genomic DNA typically consists of linear fragments, there is no need for any enzymatic digestion when performing PHAST. This also serves to eliminate any chance of encountering the "star activity" of the nucleases, where non-specific cleavage occurs under sub-optimal reaction conditions and yields erroneous TL measurement [159].

Therefore, TRF is not ideal for determining the proportion of short telomeres. To test the suitability of PHAST for this task, we prepared mixtures of short (200 bp) and relatively long (900 bp) telomeres at varying ratios. The short sub-population can be clearly distinguished in the histograms (**Figure 4.8D**). In addition, plotting the nominal fraction of 200 bp standards in each sample with the measured fraction yielded a linear fit with a gradient very close to one (**Figure 4.8E**). The small offset in the observed fraction is attributed to incomplete labeling of a small proportion of the longer telomeres. It was reported that as few as five critically short telomeres out of 92 in a metaphase cell (*i.e.* ~ 5%) is sufficient to trigger the onset of senescence, highlighting the dominant effect of the shortest telomeres on cellular phenotype [126, 160]. The ability of PHAST to differentiate between the proportion of very short telomeres in the 0% and 5% samples demonstrates its suitability for this purpose.

We also compared HeLa telomere lengths determined using PHAST and that measured with TRF [161]. TRF and similar methods cannot decouple the abundance of telomeres from their lengths. As a result, longer telomere molecules can show up on the gel as more intense, even if they are less abundant than shorter DNA molecules. Therefore, when calculating the mean telomere length from TRF gels, it is common to use the equation

$$\text{Mean Telomere Length} = \sum OD / \sum (OD/MW) \tag{16}$$

where OD is the optical density or intensity on the gel, and MW is the molecular weight of the DNA, as calculated using DNA ladders [148, 162, 163]. However, the presence of a subtelomeric region that is resistant to restriction enzyme digestion, estimated to be around 2 kb [164, 165], means that the calculated mean telomere length will underestimate the abundance of short telomeres.

To emulate TRF results, we binned the peaks detected on PHAST, and multiplied the abundance in each bin by the telomere length corresponding to the center of the bin. This yielded the intensity distribution, as might be observed on a gel, with the maximum intensity corresponding to a telomere length of 4.25 kb (**Figure 4.11, blue**).

To account for the 2 kb subtelomeric, the histogram is shifted higher by 2 kb, and the mean telomere length is back-calculated using the above equation. This gives an estimated value of 5.0 kb (**Figure 4.11, black**), which is a typically reported value. However, the actual mean telomere length is only 2.4 kb. In other words, the overestimation of average telomere length is even larger than the length of the subtelomeric region. Our estimate is also corroborated by a new electrophoresis-based method called TeSLA [166], which minimizes the length of the subtelomeric region, and gave an estimate of around 2.5 kb for HeLa cells. These results also highlight the difficulty in traditional methods for telomere measurement and argue for the need to introduce new and more convenient approaches.

4.10 Cell Numbers Needed for Telomere Sizing by PHAST

To determine the number of cells necessary to obtain a good estimate of the telomeric distribution in a sample, we used Q-Q plots to compare the distribution of varying numbers of telomeres – corresponding to around 5, 10, 20 and 50 cells – against that of approximately 100 cells (10,000 telomere-containing DNA molecules, **Figure 4.12**). Since typical q-FISH uses only around 10 metaphase spreads (corresponding to 1840 telomeric ends) to estimate the telomere distribution, we consider 100 cells more than sufficient to represent the TLs in a sample. A linear curve-fit is applied to the Q-Q plot in each case, and the residuals is used to estimate the goodness of fit.

While the linear fit is relatively good in each case, the residuals become smaller (closer to zero) as the number of cell equivalents increased from 5 to 50. We find that at ~1,900 peaks (corresponding to 20 cells, **Figure 4.11C**), the TL distribution is very similar to that of the 100-cell sample. Since this is also close to the typical number of telomeric ends in q-FISH, we collect 1,900 or more peaks for all our samples, unless otherwise stated.

4.11 Comparison between PHAST and q-FISH

In general, PNA binding to the target DNA is not only very robust, but also highly sequence-specific, as evidenced by large changes in melting temperatures resulting from single nucleotide mismatches [167, 168]. However, total genomic DNA contains a large excess of non-telomeric sequences. In a typical patient sample, telomeric DNA (~8 kb per chromosomal arm, or 0.75 Mb per cell) constitutes only around 0.01-0.02% of the total genomic DNA (~ 6 Gb). Consequently, we found that even the limited non-Watson-Crick interactions between PNA and DNA can yield an unacceptably high fluorescence background which can obscure short telomere peaks (< 300 bp) [169].

PHAST utilizes a one-tube purification protocol to remove practically all non-telomeric DNA molecules, without biasing the TL distribution (**Figures 4.5, 4.6, and 4.9**). Extracted DNA from cell lines (WI38, U2OS and R83) was purified using this method, and analyzed by PHAST. Because of the 100% mass detection efficiency, there is minimal sampling bias and a reasonably good estimate of the TL distribution is obtained with as few as 5 cell equivalents (~ 500 peaks, **Figure 4.12**). However, since the estimate improves with the number of sampled peaks, at least 1,900 peaks are collected in each sample derived from cells, corresponding to 20 genomic equivalents (or 10 metaphases). The sizing results using PHAST is compared with those from q-FISH for up to 30 cells using quantile-quantile (Q-Q) plots, which permits comparison of distributions of different units, with different number of elements (**Figure 4.8F**). For all 3 cell lines, the correlation between the PHAST and q-FISH results was largely linear, indicating that the measured telomere distributions were of the same shape for both methods.

Since q-FISH is unable to provide absolute size of telomeres, we converted the q-FISH telomere fluorescence intensity units (TFU) to telomere length by scaling the TFU to match the modal values between PHAST and q-FISH. The resulting distributions from the two methods were found to be similar. The mean TL, and fractions of short (< 0.1 times mean TL) and long telomeres (> 3 times mean TL) are tabulated (**Table 4.2**). For each cell line, the fraction of long telomeres and mean TL were essentially the same for the two methods.

Although PHAST consistently identifies fewer short peaks than q-FISH, the difference is most likely artificial. This is because q-FISH yields TFU values ranging from zero to thousands, with some of these short peaks (<2% of total) corresponding to telomere lengths shorter than the PNA probe (18 nt), based on the scaling used to normalize the distribution to PHAST. The presence of these extremely short telomeres may be due to variable detection efficiency across the imaging field in q-FISH, and suggests caution when interpreting q-FISH results for very short telomeres. It also explains the practical detection limit of 150-300 bp commonly cited for q-FISH [136], and likely accounts for the difference between the two distributions at the lower end of the range (< 400 bp), where q-FISH may underestimate the length of the short telomeres. Since a cut-off of 65 bp (corresponding to up to 3-4 bound PNA probes) is imposed for PHAST during thresholding to minimize detection of spurious peaks from the data, we placed a similar cut-off on the distribution from q-FISH, and found that the difference between the two methods was reduced (**Table 4.2**). Given the differences in the sample preparation, signal acquisition, and data handling, the agreement between the two methods is better than one might expect, and validates the PHAST method.

Clinical samples (two from cancer patients, one from a control subject) were collected and tested by both PHAST and q-FISH (**Figure 4.13**). We found that both methods yielded distributions that were not significantly different between the control and patient samples. This is not unexpected, since there is no simple relationship correlating telomere biology to cancer incidence. However, by analyzing larger patient cohorts, in a further study, and controlling for various parameters, it may be possible elucidate the role of telomere length in cancer.

4.12 Shortening of Telomeres with Population Doubling

While commercial, immortalized cell lines including those in the previous section often possess some form of telomere lengthening or maintenance mechanism, other cell types, including primary cells, experience significant telomere shortening with each cell division, eventually leading to the onset of senescence. To verify our ability to detect such changes, excised human foreskin fibroblasts were cultivated for 38 population doublings (PD), and the cells were collected at various stages. PHAST was performed on the cells at PD = 8, 20, 32 and 38, while q-FISH was performed at PD = 8, 14, 20, 26, and 32 (**Figure 4.14**). At PD = 38, the cell division happens so infrequently that no metaphase could be obtained for q-FISH. Comparing the results for the two methods, we noticed that both showed shortening trends in the telomere length. Moreover, extrapolation of the PHAST results suggests that the initial telomere length (PD = 0) was approximately 8-9 kb (**Figure 4.14C(i)**), which is a reasonable estimate corroborated by existing flow-FISH data [146]. The estimated median telomere length at PD → ∞ (i.e. at senescence) is around 2 kb, which is close to that theorized by the leading experts in this field [124, 170]. While these extrapolated estimates are not meant to be quantitative, the fact that they lie within the expected range further suggests that our telomere measurements are at least reasonably accurate.

One somewhat surprising observation was that at PD = 8 and 14, the telomere length determined using q-FISH was significantly shorter than expected, resulting in a kink in the trend. One possible explanation for this is heterogeneity of the cells in culture, where a small subpopulation of cells (potentially of a different type) divides rapidly early on and approaches replicative senescence, before a second, relatively slow-dividing subpopulation takes over. Because q-FISH analysis can only be applied to actively proliferating cells (since cells must be arrested in metaphase prior to probe hybridization) [145, 171], it may be detecting only the early subpopulation, for which the actual number of cell divisions may be far larger than the nominal PD values. On the other hand, because PHAST measures the telomere lengths of all the cells within a sample, it suffers from no such bias. This hypothesis is supported by the PHAST results, which show a secondary peak corresponding to short telomeres (Figure 4.14A(i), black arrow). As a result, the telomere length as determined by q-FISH may underestimate the telomere length for the whole population.

4.13 Application to Huntington CAG Tandem Repeats

Our method is expected to be applicable to other types of tandem repeats. To demonstrate this, we chose to test out the Huntington CAG trinucleotide repeat, one of the most well-known tandem repeats. In Huntington's Disease (HD) patients, the number of repeats range from 37 to 86, with the age of onset inversely related to the repeat length [59, 172]. For this experiment, we procured 12-nt PNA probes labeled with Alexa488 dye (PNA Bio), We also ordered an oligonucleotide that contained CAG repeats, that can bind a maximum of 16 probes (192 base long). The average repeat length measured by our method is 195 bases (**Figure 4.15**). Although the coefficient of variation is somewhat higher than that achieved for telomeres, this is attributed to the need to optimize the sample preparation process, which lies beyond the scope of this chapte.

4.14 Conclusion

This chapter has demonstrated a powerful method to determine the distribution of telomere lengths within a population of cells. Apart from routine DNA extraction, the entire sample handling process for PHAST can be completed within 2-3 hours. It addresses needs that are not met by two FISH-based methods that have emerged as leading candidates for telomere measurement, namely Flow-FISH and q-FISH.

Although Flow-FISH provides a useful clinical index for studying telomere-related disease states, the fact that it has to take into account the age of the subject suggests that the average TL information obtained is, at best, a proxy for the underlying biological trigger, most likely the accumulation of critically short telomeres. Therefore, it will fail to identify cases wherein the telomere distribution deviates from what is typical, either due to a broader distribution (**Figure 4.1A(ii)**), or catastrophic loss of telomeres (**Figure 4.1A(iii)**). The ability to test the TL in just non-adherent cells, rather than in relevant parenchymal cells [147], also limits its utility in understanding the role of telomeres in the development of many cancers.

On the other hand, q-FISH can determine not just TL distribution, but also chromosome-specific TL [145, 173]. However, chromosome-specific telomere sizing may be extraneous since the biological effects of telomeres appear to be exerted primarily by the shortest telomeres, regardless of the chromosomal arms on which they are located [133]. thermore, the requirement for live cells that can be induced into metaphase precludes using the method on archival samples or senescent cells. Even when applicable, q-FISH is still an extremely laborious technique, taking hours to days to process a single sample, which often needs to be cultured and synchronized [136, 145]. Lastly, q-FISH provides the telomere lengths in TFU, rather than actual telomere length, making comparison across experiments particularly difficult.

In contrast, the PHAST assay is capable of providing TL distribution for dividing or senescent cells; adherent or non-adherent cells; and frozen cell pellets or extracted DNA from sample banks. To our knowledge, PHAST is also unique in its ability to provide absolute telomere size distribution without needing concurrent measurement of standards, making it far more convenient. Because PHAST does not depend on any specific properties of the telomeric structures, instead taking a simple hybridization approach, it can be extended to other tandem repeats, many of which are implicated in severe neurological syndromes including Huntington's disease [172, 174] (**Chapter 4.13**). Length variations in these sequences have also been implicated in the rapid morphological changes in the domesticated dog [175]. While existing tandem repeat assays can be performed fairly routinely, they still suffer from technical and biological challenges including large allele dropout and null alleles that may be caused by primer site mutation [176, 177]. As a result, tandem repeats remain an underexplored field, with the vast majority of studies focusing on short microsatellites [178, 179]. In such instances, a PHAST-based approach can be advantageous in enabling detection without regard for the flanking sequences, and also eliminating biases arising from the amplification process. Furthermore, PHAST performs better for longer repeats (*e.g.* > 200 bp), and thus complements the existing approaches for analyzing tandem repeats.

In summary, we have demonstrated a new method that is able to determine the length of telomeres in extracted DNA rather than whole cells. By utilizing a powerful single molecule detection approach, this allows us to determine the length of individual telomere molecules, the load of critically short telomeres, and the distribution of telomere lengths in a sample, rather than merely average telomere lengths. This confers to it many of the advantages of q-FISH, but still able to work with extracted DNA and samples from frozen tissue banks.

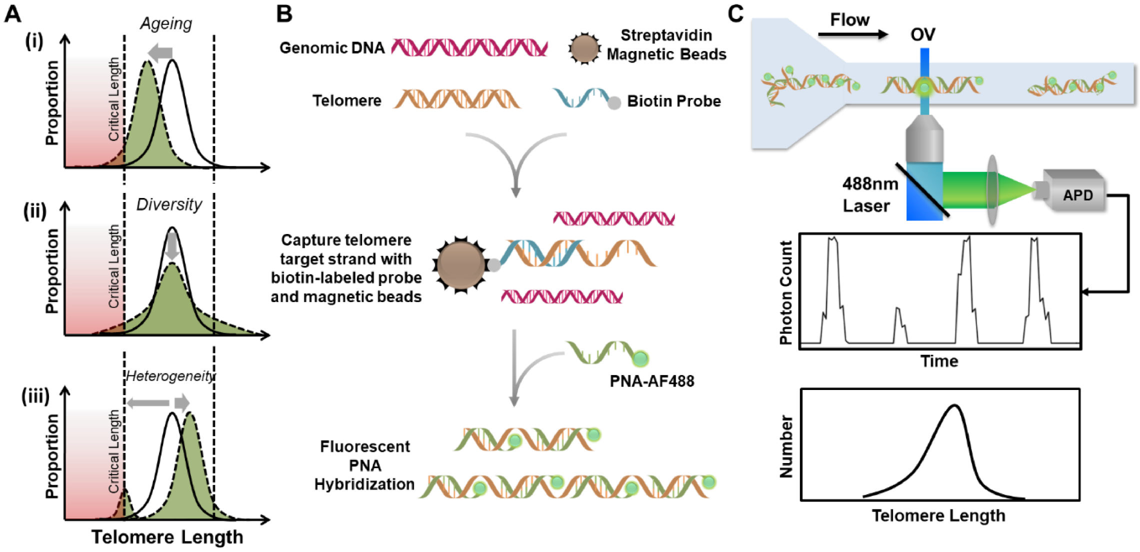
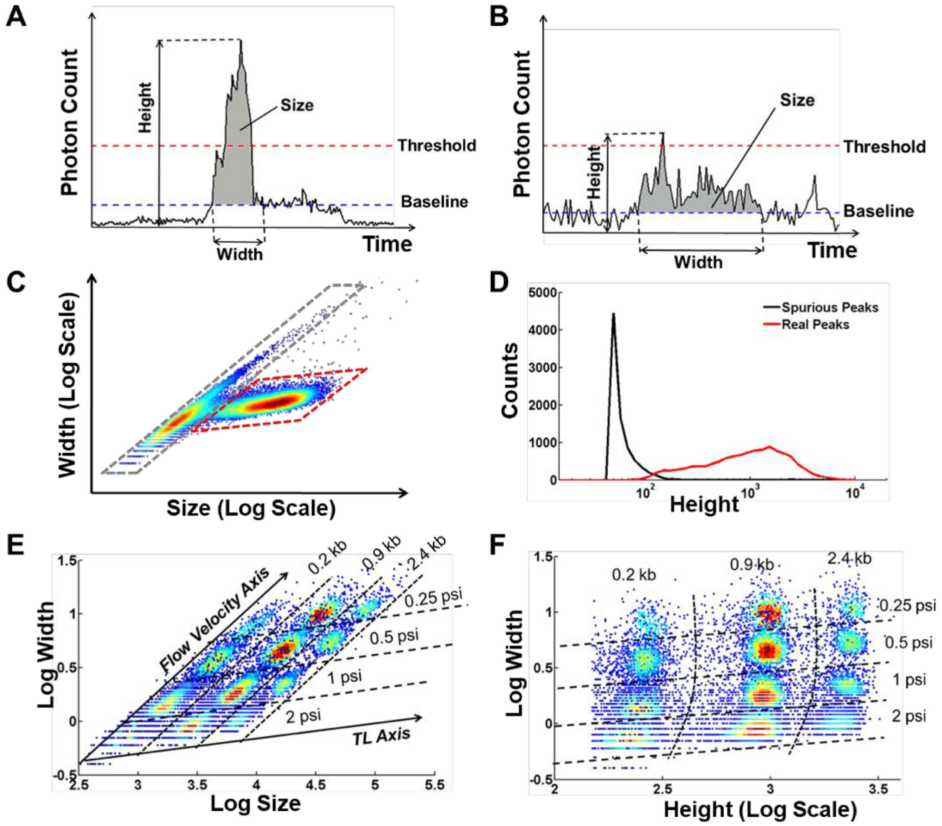

A
(i)
Ageing
Proportion
Critical Length
(ii)
Diversity
Proportion
Critical Length
(iii)
Heterogeneity
Proportion
Critical Length
Telomere Length
B
Genomic DNA
Streptavidin Magnetic Beads
Telomere
Biotin Probe
Capture telomere target strand with biotin-labeled probe and magnetic beads
PNA-AF488
Fluorescent PNA Hybridization
C
Flow
OV
488nm Laser
APD
Photon Count
Time
Number
Telomere Length

Figure 4.1. Principle of telomere measurement by PHAST

(A) The biological effects of telomeres are mediated primarily by the proportion of telomeres below a critical length. **(i)** This most commonly happens in ageing, where the telomeres in a cell population shorten with doubling. However, the proportion of short telomeres can also reach critical levels when the average telomere length is normal. **(ii)** This can happen due to the naturally occurring diversity in the shape of the distribution between individuals, or **(iii)** if the telomere distribution is highly heterogeneous. **(B)** To perform our assay, cells are lyzed to release the DNA from the nucleus and incubated with biotinylated PNA probes. Telomeric DNA is separated from genomic DNA using magnetic beads, and released after washing, whereupon fluorescent probes are hybridized to the telomeric sequences. **(C)** The labelled DNA is then flowed through a microchannel, and excited by a laser through an objective as it transits the observation volume (OV). These peaks can then be processed to yield the telomere distribution. Adapted with permission from [180]. Copyright 2019 Oxford University Press.



A
Photon Count
Height
Size
Threshold
Baseline
Width
Time
B
Photon Count
Height
Size
Threshold
Baseline
Width
Time
C
Width (Log Scale)
Size (Log Scale)
D
Counts
5000
4000
3000
2000
1000
0
Spurious Peaks
Real Peaks
10^2
10^3
10^4
Height
E
Log Width
1.5
1
0.5
0
-0.5
Flow Velocity Axis
0.2 kb
0.9 kb
2.4 kb
0.25 psi
0.5 psi
1 psi
2 psi
TL Axis
2.5
3
3.5
4
4.5
5
5.5
Log Size
F
Log Width
1.5
1
0.5
0
-0.5
0.2 kb
0.9 kb
2.4 kb
0.25 psi
0.5 psi
1 psi
2 psi
2
2.5
3
3.5
Height (Log Scale)

Figure 4.2. Analysis of PHAST data

(A) A fluorescent molecule is detected when a pre-determined threshold is exceeded. Each peak can be characterized by its *Height*, *Width* and total photon count, or *Size*, as illustrated. **(B)** Due to noise in the detector, the fluctuation in the background can occasionally exceed the threshold, whereupon a spurious peak, characterized by signal levels that hover around the baseline, and frequently large *Width* for any given *Size,* is detected. These spurious peaks can be removed by using a sufficiently high threshold, at the expense of assay sensitivity. **(C)** Alternatively, upon plotting on a log(*Size*)-log(*Width*) plot, the peaks form two distinct clusters which can be separated by gating (**Chapter 4.6**). **(D)** The real signal (red trace) can thus be distinguished from the spurious peaks (black trace), even when the shortest real peaks are shorter than the tallest spurious ones, as evidenced by the overlap between the two distributions. This cannot be achieved using a simple thresholding approach. **(E)** Although the *Size* parameter is, in principle, the most complete measure of the number of fluorophores on each telomere molecule, it is heavily dependent on the flow velocity. In fact, each *Width-Size* coordinate uniquely describes a telomere of a particular length, passing through the detection volume at a given velocity. **(F)** On the other hand, the *Height* parameter is very insensitive to flow velocity, with only a 10% decrease in the height when the velocity increases 8-fold (driving pressure scales linearly with inverse velocity). As such, the peak height is selected as the measure of telomere length. Adapted with permission from [180]. Copyright 2019 Oxford University Press.

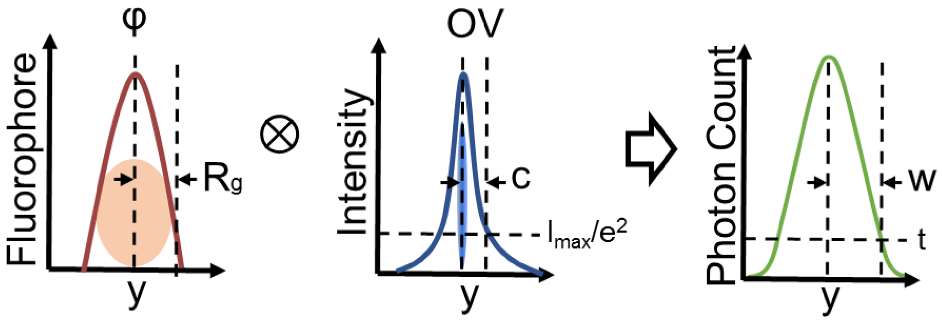
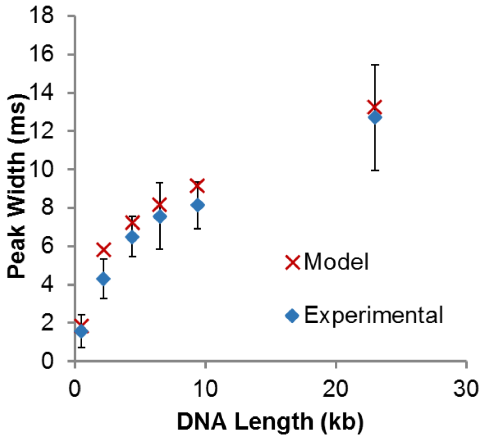
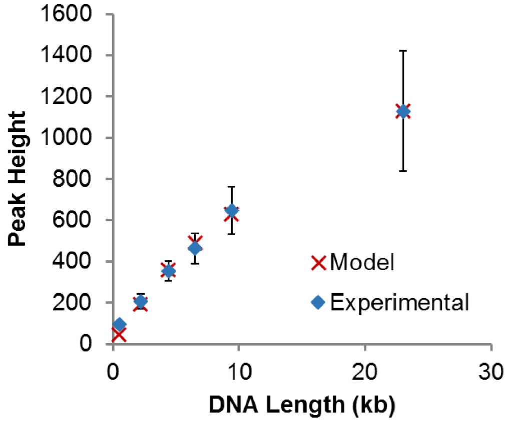

φ
Fluorophore
Rg
y
⊗
OV
Intensity
c
Imax/e2
y
Photon Count
w
t
y

Figure 4.3. Single-molecule fluorescence peak profile

The peak profile is the result of the convolution of the fluorophore distribution profile and OV profile. Depending on the relative sizes of these two contributing factors, the resulting profile can be parabolic, Gaussian or intermediate between the two in shape. Adapted with permission from [180]. Copyright 2019 Oxford University Press.



Peak Width (ms)
0
2
4
6
8
10
12
14
16
18
0
10
20
30
DNA Length (kb)
Model
Experimental



Peak Height
0
200
400
600
800
1000
1200
1400
1600
0
10
20
30
DNA Length (kb)
Model
Experimental

Figure 4.4. Comparison between the experimental and predicted peak heights and widths

Comparison between the experimental and predicted peak heights and widths based on the model (**Chapter 4.4**) demonstrates the remarkable agreement between the two. Experimental results were obtained from Hind III digested λ DNA labeled with TOTO-3. Adapted with permission from [180]. Copyright 2019 Oxford University Press.

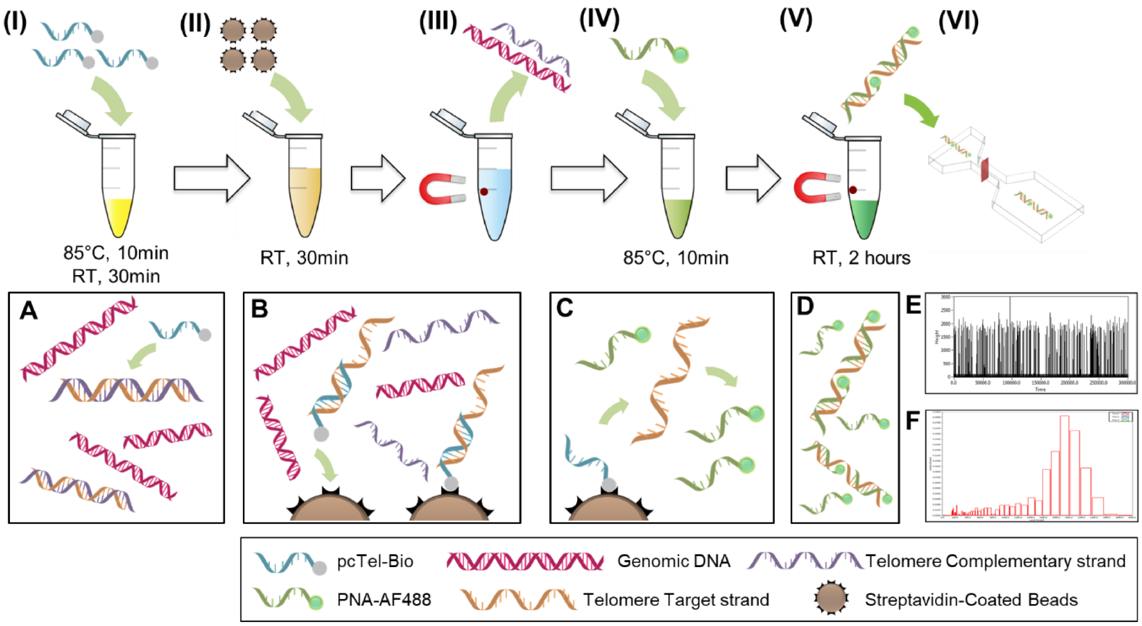

(I)
(II)
(III)
(IV)
(V)
(VI)
85°C, 10min
RT, 30min
RT, 30min
85°C, 10min
RT, 2 hours
A
B
C
D
E
F
pcTel-Bio
Genomic DNA
Telomere Complementary strand
PNA-AF488
Telomere Target strand
Streptavidin-Coated Beads

Figure 4.5. Overview of the PHAST purification protocol and performance

This figure shows the one-tube sample preparation for PHAST assay at the sampling handling (Steps (I)-(VI)) and molecular levels (Panels A-D). Telomere-containing DNA is purified by incubation with the pull-down probe, pcTel-Bio, before using streptavidin-coated magnetic beads to remove it from solution (Steps (I)-(II)). The supernatant containing genomic DNA is then removed and the beads are washed (Step (III)). At this point, the telomere-containing DNA is left on the beads (Panel B). Next, the fluorescent PNA-AF488 is introduced at 100-fold excess compared to the pcTel-Bio (Step (IV)). This ensures that even if free pcTel-Bio were present, the PNA-AF488 can successfully label the telomere DNA. Heating the sample up to 85 °C then allows the DNA to simultaneously detach from the beads, and hybridize to the fluorescent probes (Panels C-D). To minimize competition from the now free pcTel-Bio on the beads, the 2-hour incubation at room temperature is performed on a magnetic holder (Step (V)). The sample can now be column-purified and tested on CICS (Step (VI)). The purified sample yields fluorescent peaks on CICS, which can then be binned and plotted on a histogram (Panels E-F). Adapted with permission from [180]. Copyright 2019 Oxford University Press.

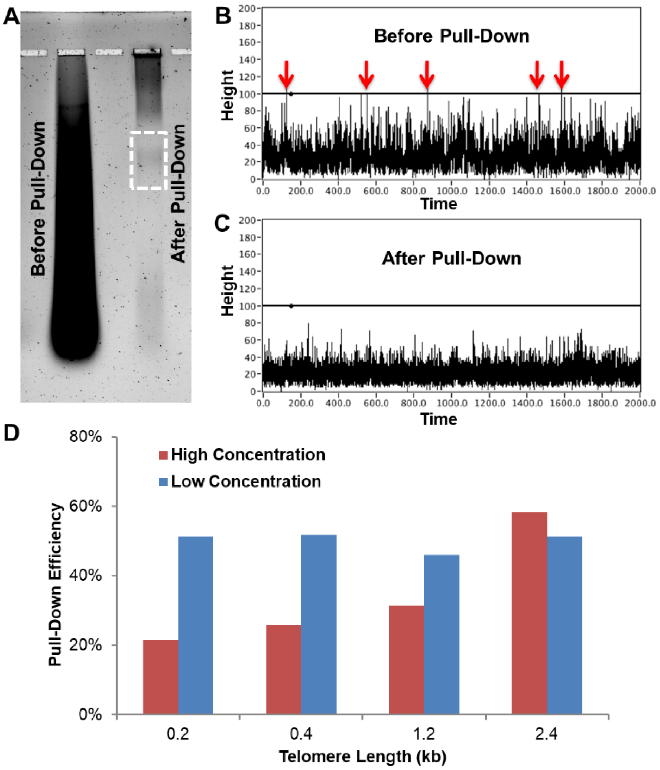

A
Before Pull-Down
After Pull-Down
B
Before Pull-Down
Height
Time
C
After Pull-Down
Height
Time
0.0
200.0
400.0
600.0
800.0
1000.0
1200.0
1400.0
1600.0
1800.0
2000.0
0
20
40
60
80
100
120
140
160
180
200
D
High Concentration
Low Concentration
Pull-Down Efficiency
80%
60%
40%
20%
0%
0.2
0.4
1.2
2.4
Telomere Length (kb)

Figure 4.6. Performance of the pull-down assay

(A) Gel electrophoresis of genomic DNA spiked with 2.4 kb telomere standards showed intense fluorescence over a broad range of DNA lengths. After pull-down, the very intense smear at the bottom of the gel is almost completely removed. The telomere standards (white dashed box) can be clearly seen. Long genomic DNA fragments were also recovered via non-specific interaction with the streptavidin-coated beads, though not in sufficient quantities to affect the signal-to-noise ratio of PHAST. **(B)** Raw signal traces for the sample before and after the pull-down assay show the effect of non-specific hybridization. Although the specificity of the PNA probes ensure that the increase in the background is manageable, the proximity of the background level to the threshold results in a significant number of peaks (red arrows) that cross the threshold. These correspond to a TL of around 65 bp, and their sheer numbers skews the resulting telomere size distribution. **(C)** After pull-down, the background returns to a level comparable to the genomic DNA-free standards. **(D)** Comparison of the pull-down at different concentrations show that at low concentrations comparable to that found naturally, the efficiency is uniform across different TLs (blue). However, when the concentration of telomeres is very high (~100-fold higher), longer telomeres are preferentially extracted. Adapted with permission from [180]. Copyright 2019 Oxford University Press.

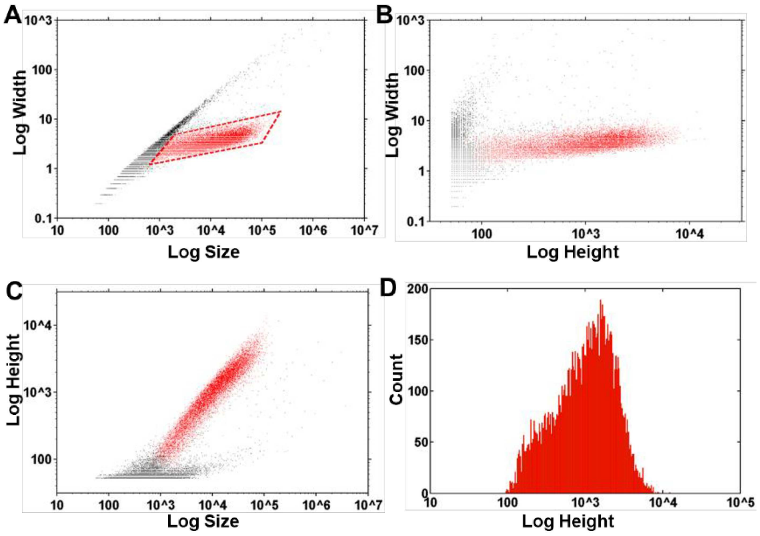

A
Log Width
10^3
100
10
1
0.1
10
100
10^3
10^4
10^5
10^6
10^7
Log Size
B
Log Width
10^3
100
10
1
0.1
100
10^3
10^4
Log Height
C
Log Height
10^4
10^3
100
10
100
10^3
10^4
10^5
10^6
10^7
Log Size
D
Count
200
150
100
50
0
10
100
10^3
10^4
10^5
Log Height

Figure 4.7. Gating strategy for PHAST

The burst sizes and widths for the peaks in a sample are plotted. The gating is performed by selecting the cluster of peaks that has a smaller slope (A). The various plots show the relationships between the different parameters (A-C). The final TL distribution can be calculated from the height distribution (D) using a simple linear formula. Adapted with permission from [180]. Copyright 2019 Oxford University Press.

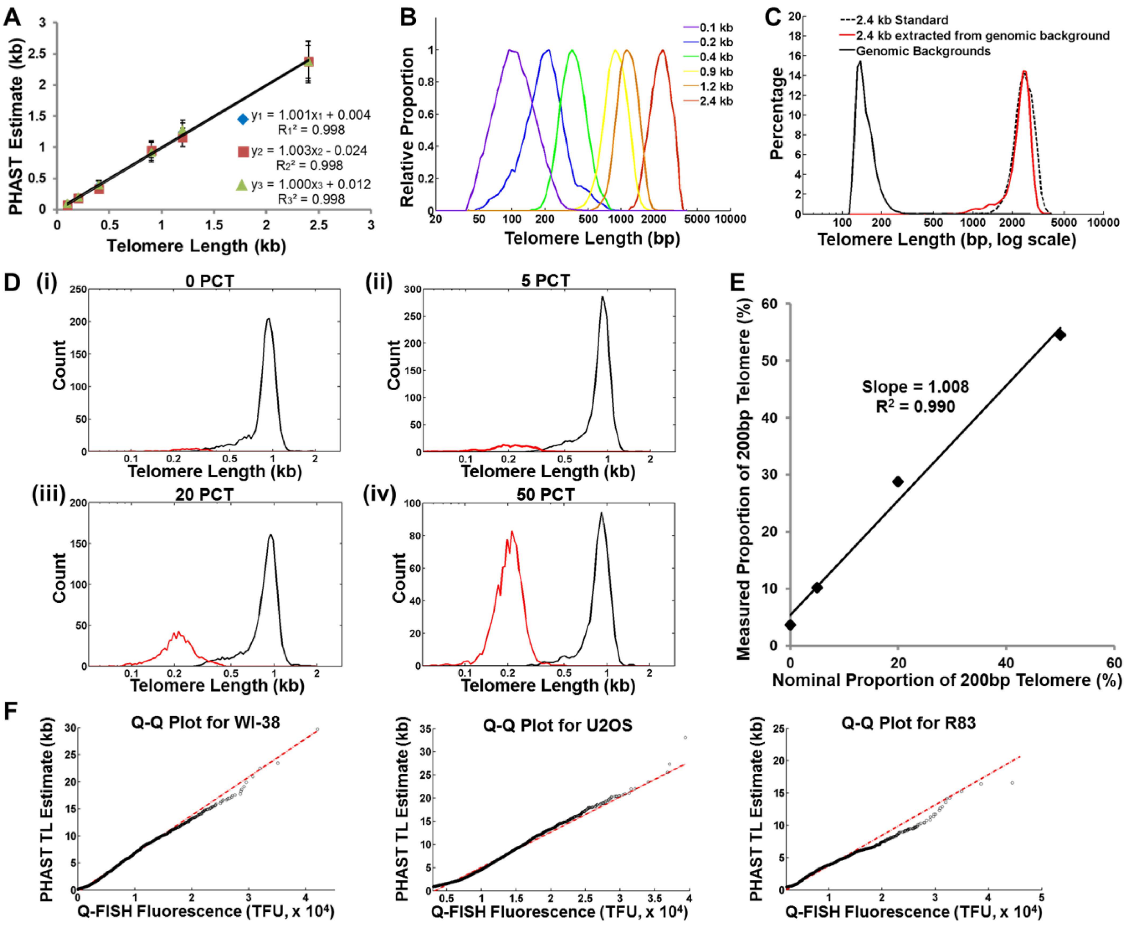
A
PHAST Estimate (kb)
Telomere Length (kb)
$y_1 = 1.001x_1 + 0.004$
$R_1^2 = 0.998$
$y_2 = 1.003x_2 - 0.024$
$R_2^2 = 0.998$
$y_3 = 1.000x_3 + 0.012$
$R_3^2 = 0.998$
B
Relative Proportion
Telomere Length (bp)
0.1 kb
0.2 kb
0.4 kb
0.9 kb
1.2 kb
2.4 kb
C
Percentage
Telomere Length (bp, log scale)
2.4 kb Standard
2.4 kb extracted from genomic background
Genomic Backgrounds
D
(i) 0 PCT
(ii) 5 PCT
(iii) 20 PCT
(iv) 50 PCT
Count
Telomere Length (kb)
E
Measured Proportion of 200bp Telomere (%)
Nominal Proportion of 200bp Telomere (%)
Slope = 1.008
$R^2 = 0.990$
F
Q-Q Plot for WI-38
Q-Q Plot for U2OS
Q-Q Plot for R83
PHAST TL Estimate (kb)
Q-FISH Fluorescence (TFU, x 10^4)

Figure 4.8. Telomere length (TL) determination using peak height

(A) Peak height is converted to TL using a simple linear equation derived from telomere standards of known lengths. The resulting three sets of TL estimates are nearly identical despite having been performed at 6-month intervals, with high linearity ($R^2 > 0.99$). This reproducibility demonstrates the ability of PHAST to provide absolute telomere lengths with minimal calibration. **(B)** The six standards can be distinguished from each other by PHAST even when pre-mixed prior to detection, and have a smaller CV as the TL increases, as evidenced by the narrowing of the distributions with longer TL. **(C)** We have demonstrated the robustness of PHAST, by detecting a 2.4 kb telomere standard with and without a large excess of genomic DNA. Because the spurious peaks far outnumber the real peaks (by a factor of 10 to 1), the traces are normalized for clarity. After gating, the sample with genomic DNA (red solid trace) was essentially the same as that without any genomic DNA (black dashed trace). **(D)** To determine the ability to detect short telomeres using PHAST, 200-bp telomere standard is mixed at varying proportions with the 900-bp standard (0-50%). **(E)** After gating and counting the two sub-populations, the measured proportion was plotted against the nominal proportion. The two numbers were very consistent, with a slightly higher proportion measured than expected. This is attributed to incomplete labeling of some of the longer telomeres, which are misidentified as short telomeres. **(F)** Q-Q plots for the telomere estimates determined by PHAST and q-FISH are presented with three different cell lines (WI-83, U2OS, and R83). This allows the distributions between two measurement methods to be compared directly even in the absence of common units and when the number of data points differs. Although PHAST and q-FISH are different methods, they yielded similar distribution shapes, as evidenced by the largely linear relationship between the corresponding quantiles. The red dashed line in each plot is the extrapolation of the interquartile range. Adapted with permission from [180]. Copyright 2019 Oxford University Press.

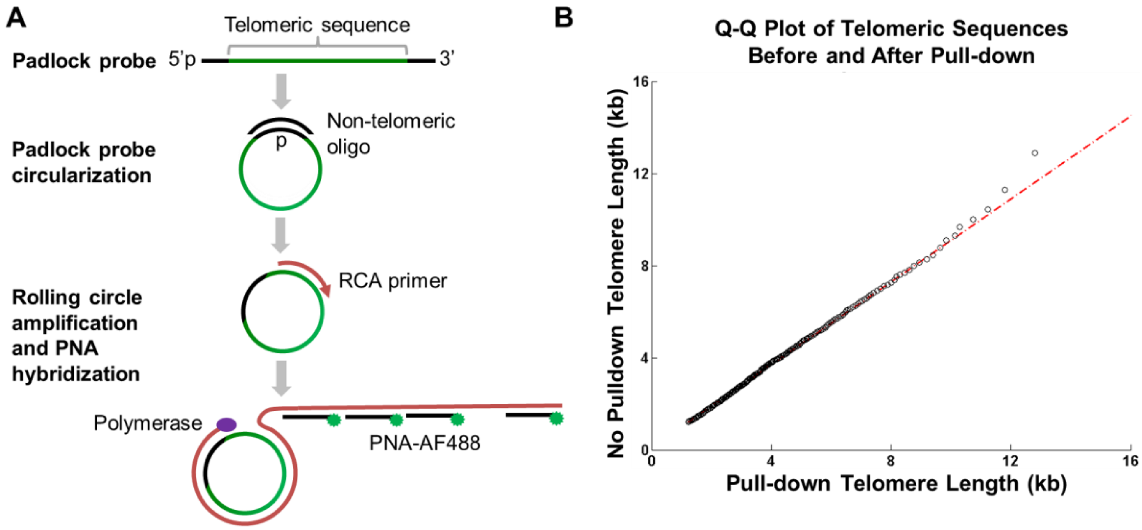

A
Telomeric sequence
Padlock probe
5'p
3'
Padlock probe circularization
Non-telomeric oligo
p
Rolling circle amplification and PNA hybridization
RCA primer
Polymerase
PNA-AF488
B
Q-Q Plot of Telomeric Sequences Before and After Pull-down
No Pulldown Telomere Length (kb)
Pull-down Telomere Length (kb)
0
4
8
12
16
0
4
8
12
16

Figure 4.9. Testing of pull-down process for long telomeres

(A) The padlock probe chemistry in conjunction with rolling circle amplification is employed to generate long DNA molecules containing telomeric sequences to simulate long telomeres for our pull-down assay. (B) The linearity of the Q-Q plot of telomere distribution determined by PHAST before and after the pull-down demonstrates that our process did not bias the telomere size distribution. Adapted with permission from [180]. Copyright 2019 Oxford University Press.

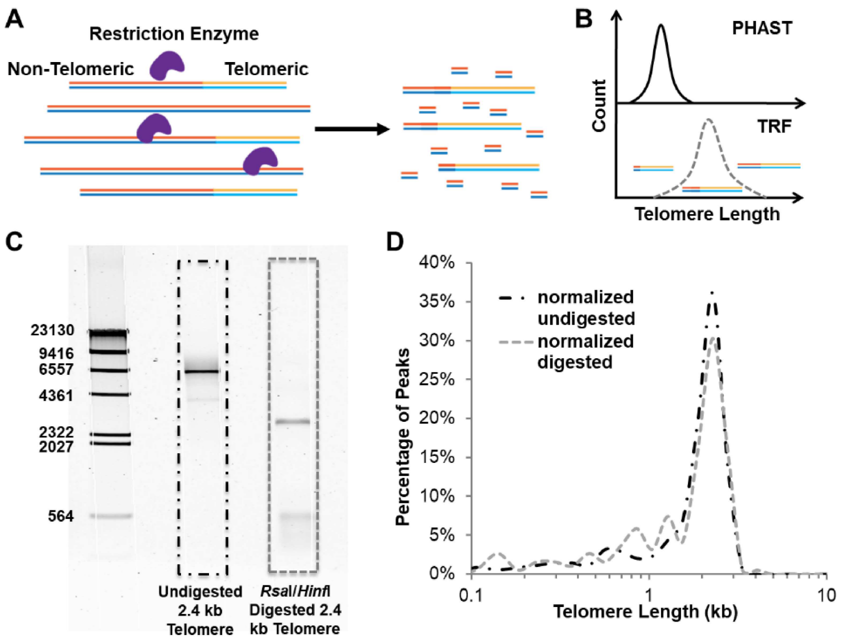

A
Restriction Enzyme
Non-Telomeric
Telomeric
B
PHAST
Count
TRF
Telomere Length
C
23130
9416
6557
4361
2322
2027
564
Undigested 2.4 kb Telomere
RsaI/*HinfI* Digested 2.4 kb Telomere
D
40%
35%
30%
25%
20%
15%
10%
5%
0%
Percentage of Peaks
normalized undigested
normalized digested
0.1
1
10
Telomere Length (kb)

Figure 4.10. Advantage of PHAST over TRF analysis by Southern blot

(**A**) Genomic DNA is digested by restriction enzymes, while telomeric DNA, which do not contain the restriction sites, remain intact. Differing lengths of sub-telomeric DNA remains on the molecule, depending on the proximity of the restriction site closest to the telomeres. (**B**) As a result, TRF overestimates the TL by varying degrees. In this schematic, telomeric DNA of identical sizes are misidentified as three molecules of different sizes, resulting in inaccurate length estimates. On the other hand, since telomere fluorescence intensity on PHAST is solely determined by the length of the telomeric sequence, the TL measurement is representative of the underlying TL distribution. (**C**) To simulate the effects of TRF on the telomere sizing, we digested 2.4 kb telomere standards in linearized plasmids with *Rsa*I/*Hinf*I restriction enzyme mix (gray dotted box). The large band shift can be clearly seen, highlighting the inability of TRF Southern blot to distinguish between the lengths of the telomeric and non-telomeric portions in a DNA molecule. In an actual sample, the shifted band will be far less well-defined, since the undigested sub-telomeric region will vary greatly in length, depending on the sequence of the DNA. (**D**) When we tested the same samples using PHAST, it is clear that the frequency (y-axis) of telomere fluorescence (x-axis) is essentially unchanged. Adapted with permission from [180]. Copyright 2019 Oxford University Press.

.

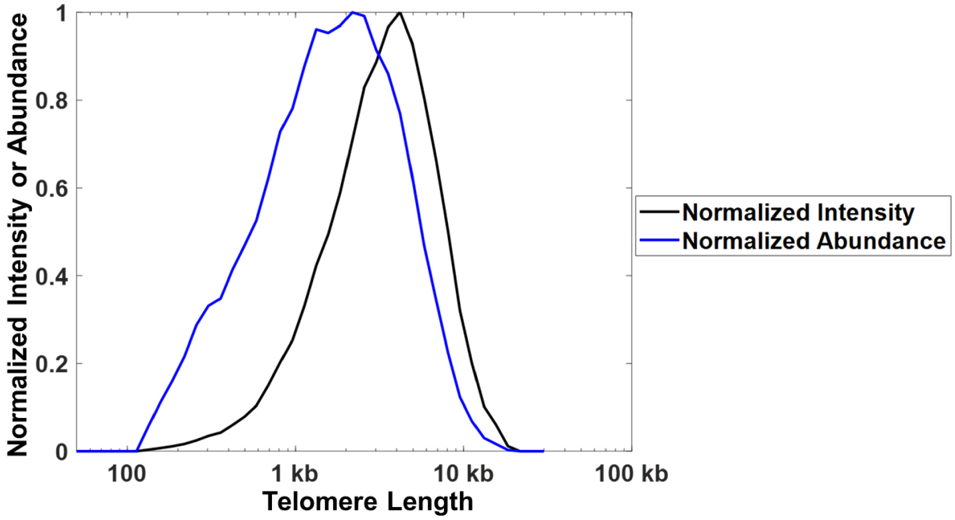

Normalized Intensity or Abundance
1
0.8
0.6
0.4
0.2
0
100
1 kb
10 kb
100 kb
Telomere Length
Normalized Intensity
Normalized Abundance

Figure 4.11. HeLa telomere lengths determined using PHAST

This figure presents the normalized abundance (blue) and intensity (black) of the telomere distribution, as determined by PHAST. The Intensity curve emulates the results from TRF, and yields a maximum at 4.25 kb. Adapted with permission from [180]. Copyright 2019 Oxford University Press.

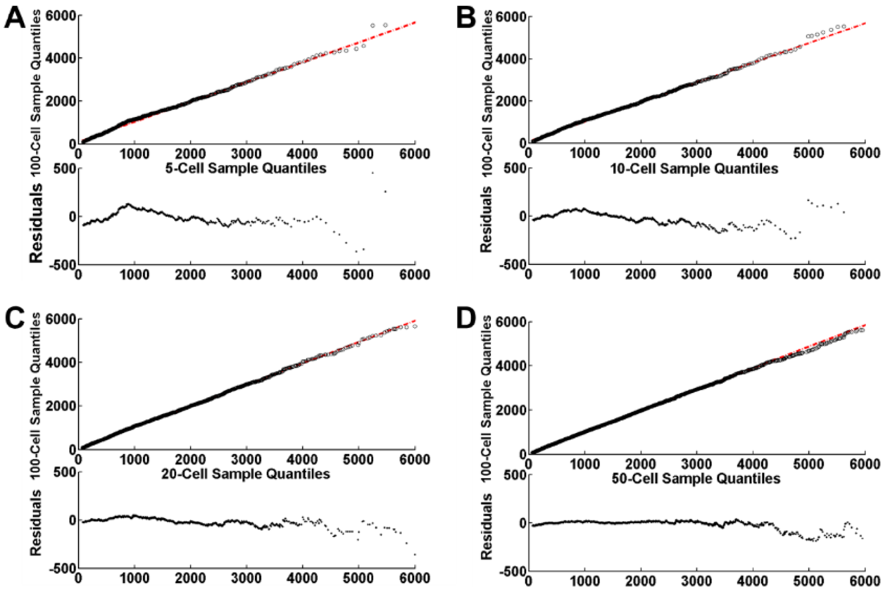

A
B
C
D
100-Cell Sample Quantiles
5-Cell Sample Quantiles
10-Cell Sample Quantiles
20-Cell Sample Quantiles
50-Cell Sample Quantiles
Residuals
6000
4000
2000
0
1000
3000
5000
500
-500

Figure 4.12. Determination of the number of telomeres needed to adequately represent the distribution

Q-Q plots are employed to compare the distribution of varying numbers of telomeres – corresponding to around 5 (A), 10 (B), 20 (C) and 50 (D) cells – against that of approximately 100 cells (10,000 telomere-containing DNA molecules). The linear fit is relatively good in each case, and the residuals become smaller (closer to zero) as the number of cell equivalents increased from 5 to 50. Adapted with permission from [180]. Copyright 2019 Oxford University Press.

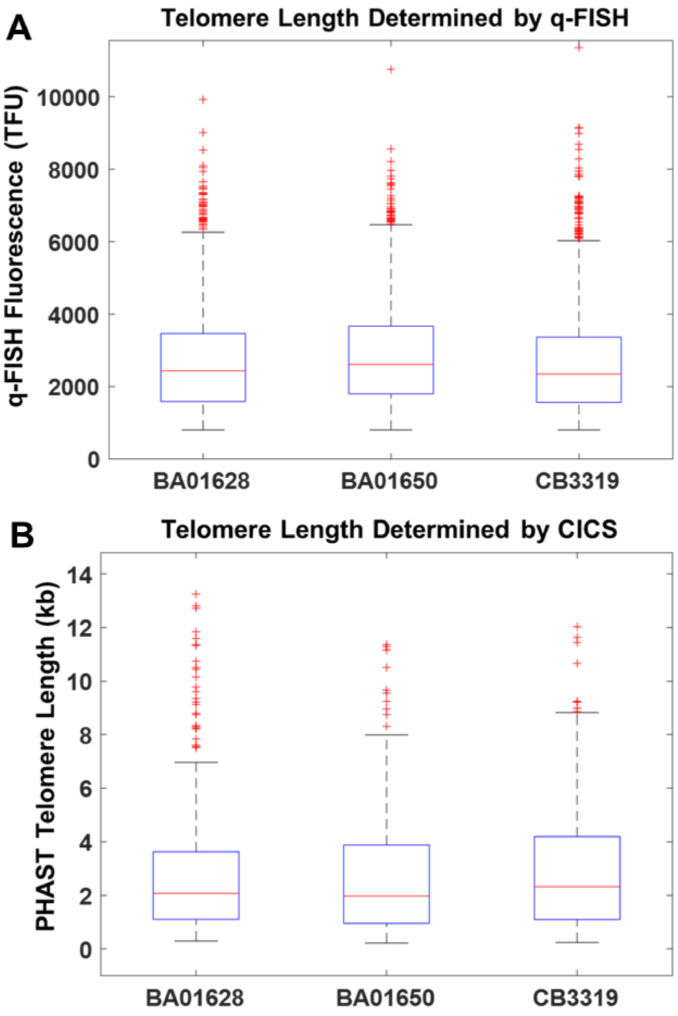

A
Telomere Length Determined by q-FISH
q-FISH Fluorescence (TFU)
0
2000
4000
6000
8000
10000
BA01628
BA01650
CB3319
B
Telomere Length Determined by CICS
PHAST Telomere Length (kb)
0
2
4
6
8
10
12
14
BA01628
BA01650
CB3319

Figure 4.13. Clinical sample tests by PHAST and q-FISH

(**A**) Q-FISH measurements do not show obvious differences in telomere distribution between the cancer patient samples (BA01628 and BA01650) and control sample (CB3319). (**B**) Similarly, PHAST measurements do not show obvious differences in telomere distribution between the cancer patient samples (BA01628 and BA01650) and control sample (CB3319). Adapted with permission from [180]. Copyright 2019 Oxford University Press.

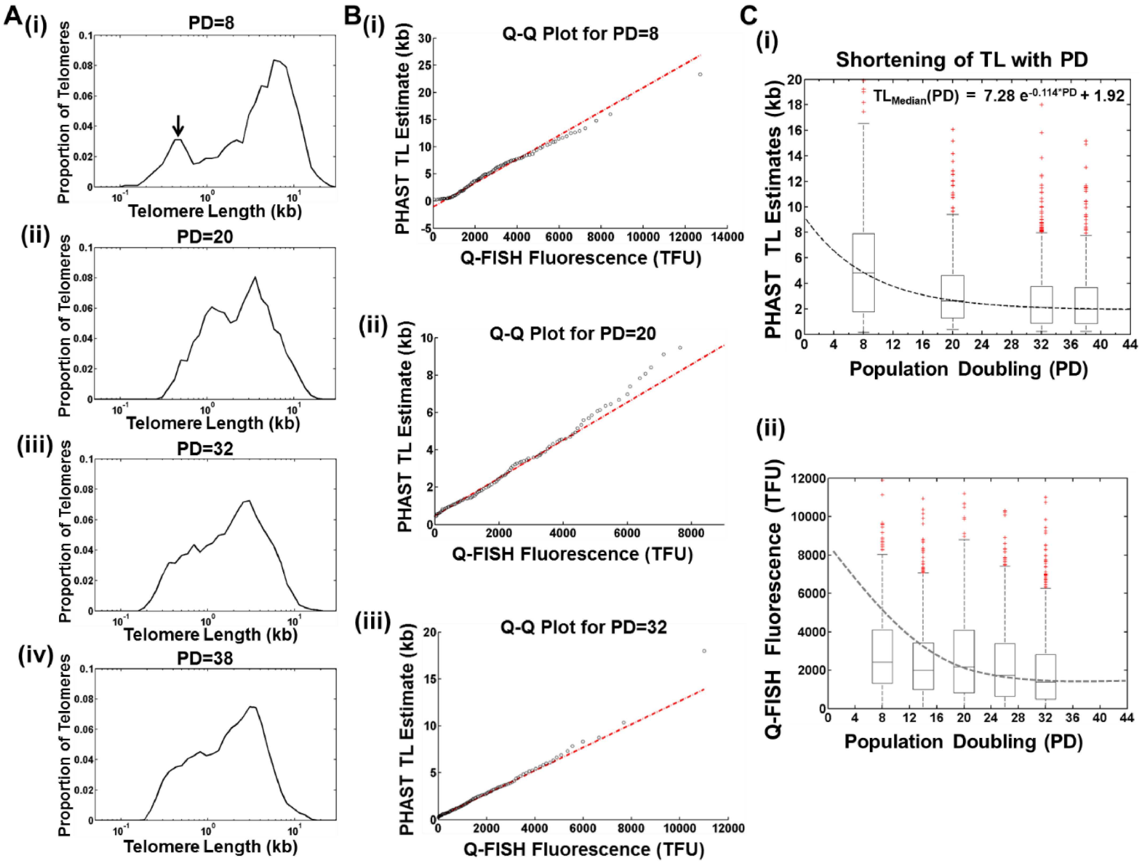

A
(i)
PD=8
(ii)
PD=20
(iii)
PD=32
(iv)
PD=38
Proportion of Telomeres
Telomere Length (kb)
0
0.02
0.04
0.06
0.08
0.1
10^{-1}
10^{0}
10^{1}
B
(i)
Q-Q Plot for PD=8
(ii)
Q-Q Plot for PD=20
(iii)
Q-Q Plot for PD=32
PHAST TL Estimate (kb)
Q-FISH Fluorescence (TFU)
C
(i)
Shortening of TL with PD
$TL_{Median}(PD) = 7.28 e^{-0.114*PD} + 1.92$
PHAST TL Estimates (kb)
Population Doubling (PD)
(ii)
Q-FISH Fluorescence (TFU)
Population Doubling (PD)
0
4
8
12
16
20
24
28
32
36
40
44

Figure 4.14. Detection of telomere shortening in primary cells with population doubling (PD) **(A)** Telomere shortening can be observed going from PD = 8, 20, 32, to 38 (i-iv). A sub-population of short telomeres was also observed at PD = 8. **(B)** The Q-Q plots for PD = 8, 20 and 32 between PHAST and q-FISH. Since these samples were normalized with the same standards for q-FISH to obtain the TFU, one would expect all three to have the same slopes. However, the slope for PD = 8 (2.2×10^{-3}) is around twice that of the other samples (1.0×10^{-3} and 1.2×10^{-3} for PD = 20 and PD = 32, respectively). This suggests that the TL determined by q-FISH is around half as long as expected from the PHAST results. **(C)** This difference is further confirmed by plotting the TL measurements against PD. (i) The median PHAST TL estimates (black dashed lines) were fitted to an exponential function. By setting the PD to 0 and ∞, we obtain TL estimates for the original patient cells and the cells at senescence, which are around 8-9 kb and 2 kb, respectively. (ii) The q-FISH results showed some degree of shortening at late passages (20-32), but stayed relatively constant at early passages (PD 8 and 14). The dashed lines show the trend an exponential decay might be expected to follow. The shorter than expected TL at low PDs is most likely the result of an artifact of the method. Adapted with permission from [180]. Copyright 2019 Oxford University Press.

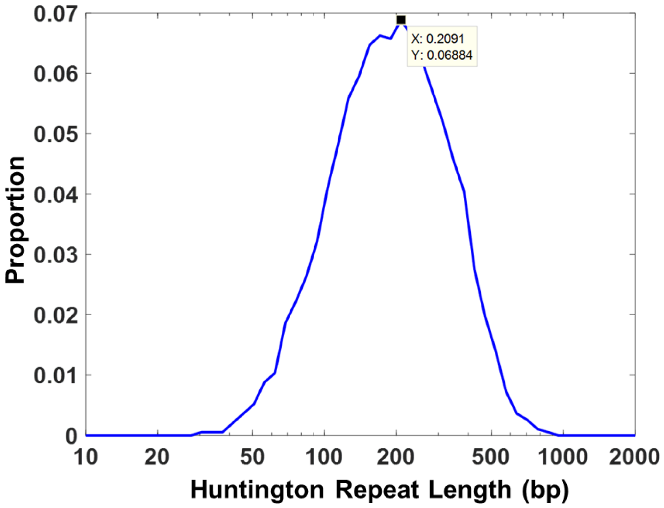

0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
Proportion
X: 0.2091
Y: 0.06884
10
20
50
100
200
500
1000
2000
Huntington Repeat Length (bp)

Figure 4.15. Huntington repeat length measurement using PHAST

Figure 4.15 Application of the method to CAG trinucleotide repeat associated with Huntington's Disease yields a good estimate. However, further optimization is needed to reduce the coefficient of variation for this sample/probe combination. Adapted with permission from [180]. Copyright 2019 Oxford University Press.

.

TL (kb)	TL Estimates Set 1 (kb)	TL Estimates Set 2 (kb)	TL Estimates Set 3 (kb)	CV of TL Estimate for Sets 1-3	TL Estimates Simultaneous Detection (kb)
0.1	0.09 ± 0.01	0.08 ± 0.01	0.10 ± 0.01	11.0%	0.08 ± 0.02
0.2	0.19 ± 0.04	0.19 ± 0.05	0.21 ± 0.03	6.8%	0.22 ± 0.03
0.4	0.39 ± 0.06	0.34 ± 0.05	0.40 ± 0.07	9.0%	0.41 ± 0.05
0.9	0.97 ± 0.12	0.96 ± 0.15	0.92 ± 0.15	3.7%	0.91 ± 0.14
1.2	1.25 ± 0.14	1.17 ± 0.14	1.28 ± 0.15	4.4%	1.29 ± 0.16
2.4	2.38 ± 0.26	2.38 ± 0.33	2.39 ± 0.31	3.7%	2.38 ± 0.26

Table 4.1. CVs for Telomere Standards on CICS

TL is calculated based on peak height using a simple linear equation derived from telomere standards of known lengths (**Chapter 4.8**). The resulting three sets of TL estimates have shown great reproducibility, which are nearly identical despite having been performed at 6-month intervals. The TL estimates has a CV of around 10-25%, with greater uncertainty for shorter telomeres. Adapted with permission from [180]. Copyright 2019 Oxford University Press.

	WI-38		**R83**		**U2OS**	
	PHAST, n = 47[a]	q-FISH, n = 32	PHAST, n = 20[a]	q-FISH, n = 19	PHAST, n = 57[a]	q-FISH, n = 30
% Short Telomeres	8.5	14.6 (10.6)[b]	2.0	4.2, (3.2)[b]	5.3	9.9 (6.9)[b]
% Long Telomeres	4.8	4.7	1.7	1.5	2.0	0.9
Mean TL (kb)	6.0	5.8	3.2	3.6	5.4	6.0

[a] *Genomic equivalents of telomere molecules detected, derived by dividing total peak count by 92, corresponding to p and q arms on each of 46 chromosomes.*

[b] *Values in parentheses represent proportions of short telomeres after removal of all data points that*

Table 4.2. Comparison between PHAST and q-FISH for telomere measurement in cell lines

Relative abundance of long and short telomeres are determined by PHAST and q-FISH, for WI-38. R83 and U2OS cell lines. For each cell line, the fraction of long telomeres and mean TL were essentially the same for the two methods. Adapted with permission from [180]. Copyright 2019 Oxford University Press.

Chapter 5

5. Conclusion

5.1 Summary

We have developed an advanced platform for multiplexed analysis of nucleic acid including a silica nanomaterial-based method for high molecular weight DNA extraction, a new multiplexed nucleic acid detection approach termed Ratiometric Fluorescence Coding and a single-molecule fluorescence spectroscopic method for measuring telomere length.

First, we have developed a novel magnetic thermoplastic silica nanomaterial named Nanobind, using a simple heat-shrinking process to create a high density of micro- and nano-scale silica lamella capable of capturing a vast amount of DNA and harbor it from fragmentation. The Nanobind extraction takes only 45 minutes using a simple bind, wash and elute process, which incorporates a tentacle binding mechanism to obtain DNA >5.7 Mb in length with binding capacities up to 1,000,000-times greater than silica paramagnetic particles. A thorough comparison against the three most commonly used DNA extracted methods (phenol-chloroform, spin column and paramagnetic particles) has been performed and demonstrated the Nanobind method enables high yields, high quality and high molecular weight DNA far surpassing that of any existing silica technology. When this DNA was used for single molecule sequencing, the high molecular weight and low DNA damage generated exceptional quality long-read libraries and lead to some of the longest reads seen to date on Pacific Biosciences' platform.

Secondly, we have developed a Ratiometric Fluorescence Coding strategy for expanding the multiplexing capacity of NAATs by encoding each nucleic acid target with a specific fluorescence ratio. Using our strategy, we have demonstrated the detection of six gene sequences for infectious diseases using only two distinct fluorophores. This is achieved by assigning each nucleic acid target with a specific number of binding sites for two probes labeled with fluorophores of distinct colors *via* padlock chemistry and RCA/HRCA reactions. By coupling our strategy with droplet-based microfluidic system, we have demonstrated our platform for multiplexed detection of gene sequences from six STDs, and the highly resolved fluorescence ratios for distinct targets offer the potential for detecting even more gene sequences using the same fluorescent probes.

Lastly, we have developed a single molecule spectroscopic method, PHAST, for determining the distribution of telomere lengths in any DNA sample by hybridizing fluorescently labeled PNA to the canonical telomere sequence, flowing the molecules through a microfluidic device, and detecting each on a confocal spectroscopic instrument one-at-a-time as fluorescence peaks. Our PHAST is analogous to flow cytometry, albeit at the molecular scale, and as in flow cytometry, each detected peak is gated according to its shape parameters to distinguish real signals from spurious ones. We have demonstrated PHAST for reproducible measurement of absolute telomere lengths rather than merely average lengths in various DNA sample including cancer cell lines, primary cells and clinical samples. This powerful single molecule detection approach also been demonstrated for determining the load of critically short telomeres and the length of individual telomere molecules with DNA from small numbers of cells.

5.2 Future Directions

So far, the commonly used approach for *de novo* sequencing is next-generation sequencing (NGS), which, however, only produces read lengths around hundreds of nucleotides. This causes substantial challenges in sequencing large complex genomes (such as repetitive sequences over

10kb) because *de novo* assemblies based on these short reads generate fragmented genome sequences and misassembled rearrangements [181]. The 3rd generation sequencing with read lengths over 10kb, such as SMRT and Nanopore technology, will improve the assembly of complex genomes but high molecular weight DNA is required to fully benefit from these technology. This thesis has demonstrated that our Nanobind extracted DNA has ultra-high molecular weight with high purity, making it perfect sample for 3rd generation sequencing. Coupling Nanobind extraction with 3rd generation sequencing will potentially enable *de novo* assembly of large complex genomes and allow access to preciously inaccessible regions of genomes. Furthermore, because the underlying polyolefin film costs pennies per square foot and the fabrication can be easily scaled using industrial roll-to-roll processes, our Nanobind is exceptionally inexpensive to produce. Considering the fact that our extraction takes a short 45 minutes process time using our own custom buffers, Nanobind can be easily manufactured as a commercial DNA extraction kit and applied for a broad array of samples including bacterial and plant cells, tissues, body fluids, and formalin-fixed paraffin embedded (FFPE) sample. In addition, since the magnetic layer on Nanobind allows automated sample processing using a robotic instrument, it can be potentially integrated onto microfluidic systems for on-chip sample preparation.

PHAST method demonstrated in the thesis for telomere length measurement provides an alternative approach for analysis of repetitive sequences in the genome. Because PHAST does not depend on any specific properties of the telomeric structures, instead taking a simple hybridization approach, it can be potentially extended to other tandem repeats, such as trinucleotide repeat sequences in Huntington or TERRA (telomeric repeat-containing RNA) [182].

In the past decades, there has been a rapid progress in multiplexed nucleic acid diagnostics for human diseases and it is clear their application will become more mainstream in the near future. In this thesis, our Ratiometric Fluorescence Coding strategy has been demonstrated to expand the multiplexed detection capacity of NAATs and can readily achieve higher multiplexing capacity by designing more fluorescent probe binding sites or using more fluorescence colors. We can easily add one more fluorescent probe with distinct color and arrange up to five probe binding sites on each padlock probes, which will potentially enable up to 40-plex detection using 3 different fluorescent probes. Moreover, since only the ratios of fluorescent probe hybridization sites are designed to vary but the sequences of the hybridization sites remain the same, the same set of fluorescent probes can be used for detecting different targets, independent of the target sequences. Therefore, we can easily expand the application of our Ratiometric Fluorescence Coding methods for other molecular diagnostics such as DNA methylation and RNA profiling. In addition, besides rolling circle amplification, our Ratiometric Fluorescence Coding can also be coupled with other nucleic acid amplification techniques, such as LCR and PCR (potentially better sensitivity), and integrated onto digital microfluidic devices or single molecule detection system for highly multiplexed nucleic acid analysis.

References

1. Watson, J.D. and F.H.C. Crick, *Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid.* Nature, 1953. **171**(4356): p. 737-738.
2. Cohen, S.N., et al., *Construction of Biologically Functional Bacterial Plasmids in-Vitro.* Proceedings of the National Academy of Sciences of the United States of America, 1973. **70**(11): p. 3240-3244.
3. Chang, A.C.Y. and S.N. Cohen, *Construction and Characterization of Amplifiable Multicopy DNA Cloning Vehicles Derived from P15a Cryptic Miniplasmid.* Journal of Bacteriology, 1978. **134**(3): p. 1141-1156.
4. Saiki, R.K., et al., *Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA-Polymerase.* Science, 1988. **239**(4839): p. 487-491.
5. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA Sequencing with Chain-Terminating Inhibitors.* Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(12): p. 5463-5467.
6. Nyren, P., B. Pettersson, and M. Uhlen, *Solid-Phase DNA Minisequencing by an Enzymatic Luminometric Inorganic Pyrophosphate Detection Assay.* Analytical Biochemistry, 1993. **208**(1): p. 171-175.
7. Cong, L., et al., *Multiplex Genome Engineering Using CRISPR/Cas Systems.* Science, 2013. **339**(6121): p. 819-823.
8. Doudna, J.A. and E. Charpentier, *The new frontier of genome engineering with CRISPR-Cas9.* Science, 2014. **346**(6213): p. 1077-+.
9. Nakayama, K., S.M. Kelly, and R. Curtiss, *Construction of an Asd+ Expression-Cloning Vector - Stable Maintenance and High-Level Expression of Cloned Genes in a Salmonella Vaccine Strain.* Bio-Technology, 1988. **6**(6): p. 693-697.
10. Bernhardt, S.L., et al., *Telomerase peptide vaccination of patients with non-resectable pancreatic cancer: a dose escalating phase I/II study.* British Journal of Cancer, 2006. **95**(11): p. 1474-1482.
11. Allison, M., *Is personalized medicine finally arriving?* Nature Biotechnology, 2008. **26**(5): p. 509-517.
12. Zander, T., et al., *Implementation of real-time genetic diagnostics and personalized treatment of non-small cell lung cancer (NSCLC) in a regional screening network.* Onkologie, 2012. **35**: p. 138-138.
13. Duan, X.L., et al., *Transgenic rice plants harboring an introduced potato proteinase inhibitor II gene are insect resistant.* Nature Biotechnology, 1996. **14**(4): p. 494-498.
14. Grange, P.A. and M.A. Mouricout, *Transferrin associated with the porcine intestinal mucosa is a receptor specific for K88ab fimbriae of Escherichia coli.* Infection and Immunity, 1996. **64**(2): p. 606-610.
15. Brem, G., *Techniques and possibilities of traceability of food: genotyping as an innovative contribution for food safety.* Deutsche Tierarztliche Wochenschrift, 2004. **111**(7): p. 273-276.
16. Deshpande, A. and P.S. White, *Multiplexed nucleic acid-based assays for molecular diagnostics of human disease.* Expert Review of Molecular Diagnostics, 2012. **12**(6): p. 645-659.
17. Wang, D., et al., *Microarray-based detection and genotyping of viral pathogens.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(24): p. 15687-15692.
18. Heller, M.J., *DNA microarray technology: Devices, systems, and applications.* Annual Review of Biomedical Engineering, 2002. **4**: p. 129-153.

19. Schoepp, N.G., et al., *Rapid pathogen-specific phenotypic antibiotic susceptibility testing using digital LAMP quantification in clinical samples.* Science Translational Medicine, 2017. **9**(410).
20. Andini, N., et al., *A "Culture" Shift: Broad Bacterial Detection, Identification, and Antimicrobial Susceptibility Testing Directly from Whole Blood.* Clinical Chemistry, 2018. **64**(10): p. 1453-1462.
21. Masek, B.J., et al., *Performance of Three Nucleic Acid Amplification Tests for Detection of Chlamydia trachomatis and Neisseria gonorrhoeae by Use of Self-Collected Vaginal Swabs Obtained via an Internet-Based Screening Program.* Journal of Clinical Microbiology, 2009. **47**(6): p. 1663-1667.
22. Park, S., et al., *Advances in microfluidic PCR for point-of-care infectious disease diagnostics.* Biotechnology Advances, 2011. **29**(6): p. 830-839.
23. Chomczynski, P. and N. Sacchi, *The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on.* Nature Protocols, 2006. **1**(2): p. 581-585.
24. Price, C.W., D.C. Leslie, and J.P. Landers, *Nucleic acid extraction techniques and application to the microchip.* Lab on a Chip, 2009. **9**(17): p. 2484-2494.
25. Gibson, D.G., et al., *Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome.* Science, 2008. **319**(5867): p. 1215-1220.
26. Hoss, M. and S. Paabo, *DNA Extraction from Pleistocene Bones by a Silica-Based Purification Method.* Nucleic Acids Research, 1993. **21**(16): p. 3913-3914.
27. Kalmar, T., et al., *A simple and efficient method for PCR amplifiable DNA extraction from ancient bones.* Nucleic Acids Research, 2000. **28**(12).
28. Ivanova, N.V., J.R. Dewaard, and P.D.N. Hebert, *An inexpensive, automation-friendly protocol for recovering high-quality DNA.* Molecular Ecology Notes, 2006. **6**(4): p. 998-1002.
29. Melzak, K.A., et al., *Driving Forces for DNA Adsorption to Silica in Perchlorate Solutions.* Journal of Colloid and Interface Science, 1996. **181**(2): p. 635-644.
30. Yang, D.Y., et al., *Technical note: Improved DNA extraction from ancient bones using silica-based spin columns.* American Journal of Physical Anthropology, 1998. **105**(4): p. 539-543.
31. Bailey, V.J., et al., *Single-Tube Analysis of DNA Methylation with Silica Superparamagnetic Beads.* Clinical Chemistry, 2010. **56**(6): p. 1022-1025.
32. Queipo-Ortuno, M.I., et al., *Comparison of seven commercial DNA extraction kits for the recovery of Brucella DNA from spiked human serum samples using real-time PCR.* European Journal of Clinical Microbiology & Infectious Diseases, 2008. **27**(2): p. 109-114.
33. Metcalf, D. and J.S. Weese, *Evaluation of commercial kits for extraction of DNA and RNA from Clostridium difficile.* Anaerobe, 2012. **18**(6): p. 608-613.
34. Quail, M.A., et al., *A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.* BMC Genomics, 2012. **13**.
35. Liu, L., et al., *Comparison of Next-Generation Sequencing Systems.* Journal of Biomedicine and Biotechnology, 2012.
36. Chin, C.S., et al., *Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data.* Nature Methods, 2013. **10**(6): p. 563-+.
37. Ghanem, A., R. Healey, and F.G. Adly, *Current trends in separation of plasmid DNA vaccines: A review.* Analytica Chimica Acta, 2013. **760**: p. 1-15.
38. Urthaler, J., et al., *Industrial Manufacturing of Plasmid-DNA Products for Gene Vaccination and Therapy.* Gene Vaccines, 2012: p. 311-330.

39. Prazeres, D.M.F., et al., *Large-scale production of pharmaceutical-grade plasmid DNA for gene therapy: problems and bottlenecks.* Trends in Biotechnology, 1999. **17**(4): p. 169-174.
40. Cunningham, J.M., et al., *Telomere Length Varies By DNA Extraction Method: Implications for Epidemiologic Research.* Cancer Epidemiology Biomarkers & Prevention, 2013. **22**(11): p. 2047-2054.
41. Roberts, R.J., M.O. Carneiro, and M.C. Schatz, *The advantages of SMRT sequencing.* Genome Biology, 2013. **14**(7).
42. Gaydos, C.A., et al., *Comparison of three nucleic acid amplification tests for detection of Chlamydia trachomatis in urine specimens.* Journal of Clinical Microbiology, 2004. **42**(7): p. 3041-3045.
43. Moreno, J.G., et al., *Detection of Hematogenous Micrometastasis in Patients with Prostate-Cancer.* Cancer Research, 1992. **52**(21): p. 6110-6112.
44. Bernard, P.S. and C.T. Wittwer, *Real-time PCR technology for cancer diagnostics.* Clinical Chemistry, 2002. **48**(8): p. 1178-1185.
45. Higuchi, R., et al., *Simultaneous Amplification and Detection of Specific DNA-Sequences.* Bio-Technology, 1992. **10**(4): p. 413-417.
46. Wittwer, C.T., et al., *Continuous fluorescence monitoring of rapid cycle DNA amplification.* Biotechniques, 1997. **22**(1): p. 130-&.
47. Pongers-Willemse, M.J., et al., *Real-time quantitative PCR for the detection of minimal residual disease in acute lymphoblastic leukemia using junctional region specific TaqMan probes.* Leukemia, 1998. **12**(12): p. 2006-2014.
48. Weller, S.A., et al., *Detection of Ralstonia solanacearum strains with a quantitative, multiplex, real-time, fluorogenic PCR (TaqMan) assay.* Applied and Environmental Microbiology, 2000. **66**(7): p. 2853-2858.
49. Schouten, J.P., et al., *Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification.* Nucleic Acids Research, 2002. **30**(12).
50. Chung, B., et al., *Multiplex quantitative foodborne pathogen detection using high resolution CE-SSCP coupled stuffer-free multiplex ligation-dependent probe amplification.* Electrophoresis, 2012. **33**(9-10): p. 1477-1481.
51. Elenitoba-Johnson, K.S.J., et al., *Multiplex PCR by multicolor fluorimetry and fluorescence melting curve analysis.* Nature Medicine, 2001. **7**(2): p. 249-253.
52. Liao, Y.Q., et al., *Combination of fluorescence color and melting temperature as a two-dimensional label for homogeneous multiplex PCR detection.* Nucleic Acids Research, 2013. **41**(7).
53. Johnson-Buck, A., et al., *A guide to nucleic acid detection by single-molecule kinetic fingerprinting.* Methods, 2019. **153**: p. 3-12.
54. Potapov, V. and J.L. Ong, *Examining Sources of Error in PCR by Single-Molecule Sequencing.* Plos One, 2017. **12**(1).
55. Dabney, J. and M. Meyer, *Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries.* Biotechniques, 2012. **52**(2): p. 87-+.
56. Kebschull, J.M. and A.M. Zador, *Sources of PCR-induced distortions in high-throughput sequencing data sets.* Nucleic Acids Research, 2015. **43**(21).
57. Walsh, K.M., et al., *Variants near TERT and TERC influencing telomere length are associated with high-grade glioma risk.* Nat Genet, 2014. **advance online publication**.
58. Codd, V., et al., *Identification of seven loci affecting mean telomere length and their association with disease.* Nat Genet, 2013. **45**(4): p. 422-427.
59. Duyao, M., et al., *Trinucleotide repeat length instability and age of onset in Huntington's disease.* Nature genetics, 1993. **4**(4): p. 387.

60. Stevens, A.J., et al., *Allelic Dropout During Polymerase Chain Reaction due to G-Quadruplex Structures and DNA Methylation Is Widespread at Imprinted Human Loci.* G3-Genes Genomes Genetics, 2017. **7**(3): p. 1019-1025.
61. Castro, A. and J.G.K. Williams, *Single-molecule detection of specific nucleic acid sequences in unamplified genomic DNA.* Analytical Chemistry, 1997. **69**(19): p. 3915-3920.
62. Chao, S.Y., et al., *Quantification of low concentrations of DNA using single molecule detection and velocity measurement in a microchannel.* Journal of Fluorescence, 2007. **17**(6): p. 767-774.
63. Nie, S.M. and R.N. Zare, *Optical detection of single molecules.* Annual Review of Biophysics and Biomolecular Structure, 1997. **26**: p. 567-596.
64. Weiss, S., *Fluorescence spectroscopy of single biomolecules.* Science, 1999. **283**(5408): p. 1676-1683.
65. Moerner, W.E. and D.P. Fromm, *Methods of single-molecule fluorescence spectroscopy and microscopy.* Review of Scientific Instruments, 2003. **74**(8): p. 3597-3619.
66. Zhang, C.Y., et al., *Single-quantum-dot-based DNA nanosensor.* Nature Materials, 2005. **4**(11): p. 826-831.
67. Friedrich, S.M., K.J. Liu, and T.H. Wang, *Single Molecule Hydrodynamic Separation Allows Sensitive and Quantitative Analysis of DNA Conformation and Binding Interactions in Free Solution.* J Am Chem Soc, 2016. **138**(1): p. 319-27.
68. Liu, K.J. and T.-H. Wang, *Cylindrical Illumination Confocal Spectroscopy: Rectifying the Limitations of Confocal Single Molecule Spectroscopy through One-Dimensional Beam Shaping.* Biophysical journal, 2008. **95**(6): p. 2964-2975.
69. Liu, K.J., et al., *Decoding circulating nucleic acids in human serum using microfluidic single molecule spectroscopy.* Journal of the American Chemical Society, 2010. **132**(16): p. 5793-5798.
70. Koren, S. and A.M. Phillippy, *One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly.* Curr Opin Microbiol, 2015. **23**: p. 110-20.
71. Zheng, G.X., et al., *Haplotyping germline and cancer genomes with high-throughput linked-read sequencing.* Nature biotechnology, 2016. **34**(3): p. 303-11.
72. Lin, J., et al., *Whole-genome shotgun optical mapping of Deinococcus radiodurans.* Science, 1999. **285**(5433): p. 1558-62.
73. Lam, E.T., et al., *Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly.* Nat Biotechnol, 2012. **30**(8): p. 771-6.
74. Boom, R., et al., *Rapid and Simple Method for Purification of Nucleic-Acids.* Journal of Clinical Microbiology, 1990. **28**(3): p. 495-503.
75. Vogelstein, B. and D. Gillespie, *Preparative and Analytical Purification of DNA from Agarose.* Proceedings of the National Academy of Sciences of the United States of America, 1979. **76**(2): p. 615-619.
76. Vandeventer, P.E., et al., *Multiphasic DNA adsorption to silica surfaces under varying buffer, pH, and ionic strength conditions.* J Phys Chem B, 2012. **116**(19): p. 5661-70.
77. Poeckh, T., et al., *Adsorption and elution characteristics of nucleic acids on silica surfaces and their use in designing a miniaturized purification unit.* Anal Biochem, 2008. **373**(2): p. 253-62.
78. Tian, H., A.F. Huhmer, and J.P. Landers, *Evaluation of silica resins for direct and efficient extraction of DNA from complex biological matrices in a miniaturized format.* Anal Biochem, 2000. **283**(2): p. 175-91.
79. Hegab, H.M., S.M. Ebrahim, and M. Op de Beeck, *In-Flow DNA Extraction Using On-Chip Microfluidic Amino-Coated Silicon Micropillar Array Filter.* Journal of Biosensors & Bioelectronics, 2013. **4**(4).

80. Solberg, S.M. and C.C. Landry, *Adsorption of DNA into mesoporous silica.* Journal of Physical Chemistry B, 2006. **110**(31): p. 15261-15268.
81. Zhang, M.P., et al., *Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research.* Nature Protocols, 2012. **7**(3): p. 467-478.
82. Brunauer, S., P.H. Emmett, and E. Teller, *Adsorption of gases in multimolecular layers.* Journal of the American Chemical Society, 1938. **60**: p. 309-319.
83. Huang, F. and H. Liang, *Adsorption behaviors of DNA/cation complexes on amino and silica chip surfaces: a dual polarization interferometry study.* ACS Appl Mater Interfaces, 2013. **5**(11): p. 5025-33.
84. Nguyen, T.H. and M. Elimelech, *Plasmid DNA adsorption on silica: kinetics and conformational changes in monovalent and divalent salts.* Biomacromolecules, 2007. **8**(1): p. 24-32.
85. Liu, K.J., et al., *Single-molecule analysis enables free solution hydrodynamic separation using yoctomole levels of DNA.* J Am Chem Soc, 2011. **133**(18): p. 6898-901.
86. Lander, E.S. and M.S. Waterman, *Genomic mapping by fingerprinting random clones: a mathematical analysis.* Genomics, 1988. **2**(3): p. 231-9.
87. Efimenko, K., et al., *Nested self-similar wrinkling patterns in skins.* Nature Materials, 2005. **4**(4): p. 293-297.
88. Sharma, H., et al., *Unconventional Low-Cost Fabrication and Patterning Techniques for Point of Care Diagnostics.* Annals of Biomedical Engineering, 2011. **39**(4): p. 1313-1327.
89. Zhao, X.M., et al., *Fabrication of microstructures using shrinkable polystyrene films.* Sensors and Actuators a-Physical, 1998. **65**(2-3): p. 209-217.
90. Kim, Y.H., et al., *Hierarchical Nanoflake Surface Driven by Spontaneous Wrinkling of Polyelectrolyte/Metal Complexed Films.* ACS Nano, 2012. **6**(2): p. 1082-1093.
91. Zhang, Y., et al., *A Simple Thermoplastic Substrate Containing Hierarchical Silica Lamellae for High-Molecular-Weight DNA Extraction.* Advanced Materials, 2016. **28**(48): p. 10630-+.
92. Boehm, C.D., *Use of Polymerase Chain-Reaction for Diagnosis of Inherited Disorders.* Clinical Chemistry, 1989. **35**(9): p. 1843-1848.
93. He, L.P., et al., *Detection and quantification of mitochondrial DNA deletions in individual cells by real-time PCR.* Nucleic Acids Research, 2002. **30**(14).
94. Monis, P.T. and S. Giglio, *Nucleic acid amplification-based techniques for pathogen detection and identification.* Infection Genetics and Evolution, 2006. **6**(1): p. 2-12.
95. Zhao, Y., et al., *Engineered Janus probes modulate nucleic acid amplification to expand the dynamic range for direct detection of viral genomes in one microliter crude serum samples.* Chemical Science, 2018. **9**(2): p. 392-397.
96. Landegren, U., et al., *A Ligase-Mediated Gene Detection Technique.* Science, 1988. **241**(4869): p. 1077-1080.
97. Abravaya, K., et al., *Detection of Point Mutations with a Modified Ligase Chain-Reaction (Gap-Lcr).* Nucleic Acids Research, 1995. **23**(4): p. 675-682.
98. Compton, J., *Nucleic-Acid Sequence-Based Amplification.* Nature, 1991. **350**(6313): p. 91-92.
99. Heim, A., et al., *Highly sensitive detection of gene expression of an intronless gene: amplification of mRNA, but not genomic DNA by nucleic acid sequence based amplification (NASBA).* Nucleic Acids Research, 1998. **26**(9): p. 2250-2251.
100. Eriksson, R., et al., *Multiplex and quantifiable detection of nucleic acid from pathogenic fungi using padlock probes, generic real time PCR and specific suspension array readout.* Journal of Microbiological Methods, 2009. **78**(2): p. 195-202.

101. Kilburn, D., et al., *A Multiplex Ligation Assay for miRNA Copy Number Profiling.* Microrna Profiling: Methods and Protocols, 2017. **1509**: p. 185-193.
102. Song, Y.K., et al., *Determination of absolute expression profiles using multiplexed miRNA analysis.* Plos One, 2017. **12**(7).
103. Huang, Q.Y., et al., *Multicolor Combinatorial Probe Coding for Real-Time PCR.* Plos One, 2011. **6**(1).
104. Huang, Q.Y., Q.H. Hu, and Q.G. Li, *Identification of 8 foodborne pathogens by multicolor combinational probe coding technology in a single real-time PCR.* Clinical Chemistry, 2007. **53**(10): p. 1741-1748.
105. Rajagopal, A., et al., *Supercolor coding methods for large-scale multiplexing of biochemical assays.* Anal Chem, 2013. **85**(16): p. 7629-36.
106. Liu, D.Y., et al., *Rolling circle DNA synthesis: Small circular oligonucleotides as efficient templates for DNA polymerases.* Journal of the American Chemical Society, 1996. **118**(7): p. 1587-1594.
107. Lizardi, P.M., et al., *Mutation detection and single-molecule counting using isothermal rolling-circle amplification.* Nature Genetics, 1998. **19**(3): p. 225-232.
108. Schweitzer, B., et al., *Immunoassays with rolling circle DNA amplification: A versatile platform for ultrasensitive antigen detection.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(18): p. 10113-10119.
109. Zhao, Y.X., et al., *Isothermal Amplification of Nucleic Acids.* Chemical Reviews, 2015. **115**(22): p. 12491-12545.
110. Dean, F.B., et al., *Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification.* Genome Research, 2001. **11**(6): p. 1095-1099.
111. Nielsen, P.E. and M. Egholm, *An introduction to peptide nucleic acid.* Curr Issues Mol Biol, 1999. **1**(1-2): p. 89-104.
112. Demidov, V.V., et al., *Kinetics and Mechanism of Polyamide (Peptide) Nucleic-Acid Binding to Duplex DNA.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(7): p. 2637-2641.
113. Ali, M.M., et al., *Rolling circle amplification: a versatile tool for chemical biology, materials science and medicine.* Chemical Society Reviews, 2014. **43**(10): p. 3324-3341.
114. Demidov, V.V., *Rolling-circle amplification in DNA diagnostics: the power of simplicity.* Expert Rev Mol Diagn, 2002. **2**(6): p. 542-8.
115. Berlier, J.E., et al., *Quantitative comparison of long-wavelength Alexa Fluor dyes to Cy dyes: fluorescence of the dyes and their bioconjugates.* J Histochem Cytochem, 2003. **51**(12): p. 1699-712.
116. Anderson, G.P. and N.L. Nerurkar, *Improved fluoroimmunoassays using the dye Alexa Fluor 647 with the RAPTOR, a fiber optic biosensor.* Journal of Immunological Methods, 2002. **271**(1-2): p. 17-24.
117. Hsieh, C.C. and P.S. Doyle, *Studying confined polymers using single-molecule DNA experiments.* Korea-Australia Rheology Journal, 2008. **20**(3): p. 127-142.
118. Araki, S., et al., *Hydrodynamic radius of circular DNA is larger than that of linear DNA.* Chemical Physics Letters, 2006. **418**(1-3): p. 255-259.
119. Wang, J., et al., *Peptide nucleic acid probes for sequence-specific DNA biosensors.* Journal of the American Chemical Society, 1996. **118**(33): p. 7667-7670.
120. Igloi, G.L., *Variability in the stability of DNA-peptide nucleic acid (PNA) single-base mismatched duplexes: Real-time hybridization during affinity electrophoresis in PNA-containing gels.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(15): p. 8562-8567.
121. Li, H.T., et al., *Ultrasensitive coincidence fluorescence detection of single DNA molecules.* Analytical Chemistry, 2003. **75**(7): p. 1664-1670.

122. Liu, K.J. and T.H. Wang, *Cylindrical illumination confocal spectroscopy: Rectifying the limitations of confocal single molecule spectroscopy through one-dimensional beam shaping.* Biophysical Journal, 2008. **95**(6): p. 2964-2975.
123. Zhang, Y., et al., *Ratiometric Fluorescence Coding for Multiplex Nucleic Acid Amplification Testing.* Analytical Chemistry, 2018. **90**(20): p. 12180-12186.
124. Levy, M.Z., et al., *Telomere end-replication problem and cell aging.* Journal of molecular biology, 1992. **225**(4): p. 951-960.
125. Sarek, G., et al., *Molecular basis of telomere dysfunction in human genetic diseases.* Nature Structural & Molecular Biology, 2015. **22**(11): p. 867-874.
126. Armanios, M. and E.H. Blackburn, *The telomere syndromes.* Nature Reviews Genetics, 2012. **13**(10): p. 693-704.
127. Oeseburg, H., et al., *Telomere biology in healthy aging and disease.* Pflügers Archiv-European Journal of Physiology, 2010. **459**(2): p. 259-268.
128. Murnane, J.P. and L. Sabatier, *Chromosome rearrangements resulting from telomere dysfunction and their role in cancer.* Bioessays, 2004. **26**(11): p. 1164-1174.
129. Thompson, S.L. and D.A. Compton, *Chromosomes and cancer cells.* Chromosome research, 2011. **19**(3): p. 433-444.
130. Xu, L., S. Li, and B.A. Stohr, *The role of telomere biology in cancer.* Annual Review of Pathology: Mechanisms of Disease, 2013. **8**: p. 49-78.
131. Lansdorp, P.M., et al., *Heterogeneity in telomere length of human chromosomes.* Human molecular genetics, 1996. **5**(5): p. 685-691.
132. Martens, U.M., et al., *Short telomeres on human chromosome 17p.* Nat Genet, 1998. **18**(1): p. 76-80.
133. Hemann, M., et al., *The Shortest Telomere, Not Average Telomere Length, Is Critical for Cell Viability and Chromosome Stability.* Cell, 2001. **107**(1): p. 67-77.
134. Liu, Y., et al., *Preferential maintenance of critically short telomeres in mammalian cells heterozygous for mTert.* Proceedings of the National Academy of Sciences, 2002. **99**(6): p. 3597-3602.
135. Bendix, L., et al., *The load of short telomeres, estimated by a new method, Universal STELA, correlates with number of senescent cells.* Aging cell, 2010. **9**(3): p. 383-397.
136. Vera, E. and M.A. Blasco, *Beyond average: potential for measurement of short telomeres.* Aging (Albany NY), 2012. **4**(6): p. 379.
137. Capper, R., et al., *The nature of telomere fusion and a definition of the critical telomere length in human cells.* Genes & development, 2007. **21**(19): p. 2495-2508.
138. Abdallah, P., et al., *A two-step model for senescence triggered by a single critically short telomere.* Nature Cell Biology, 2009. **11**(8): p. 988-993.
139. Gu, P., et al., *CTC1 deletion results in defective telomere replication, leading to catastrophic telomere loss and stem cell exhaustion.* The EMBO journal, 2012. **31**(10): p. 2309-2321.
140. Lustig, A.J., *Clues to catastrophic telomere loss in mammals from yeast telomere rapid deletion.* Nature Reviews Genetics, 2003. **4**(11): p. 916-923.
141. Bender, H.S., et al., *Extreme telomere length dimorphism in the tasmanian devil and related marsupials suggests parental control of telomere length.* PloS one, 2012. **7**(9): p. e46195.
142. Ramunas, J., et al., *Transient delivery of modified mRNA encoding TERT rapidly extends telomeres in human cells.* The FASEB Journal, 2015.
143. de Jesus, B.B., et al., *Telomerase gene therapy in adult and old mice delays aging and increases longevity without increasing cancer.* EMBO molecular medicine, 2012. **4**(8): p. 691-704.
144. Townsley, D.M., et al., *Danazol Treatment for Telomere Diseases.* New England Journal of Medicine, 2016. **374**(20): p. 1922-1931.

145. Aubert, G., M. Hills, and P.M. Lansdorp, *Telomere length measurement—Caveats and a critical assessment of the available technologies and tools.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2012. **730**(1–2): p. 59-67.
146. Aubert, G. and P.M. Lansdorp, *Telomeres and aging.* Physiological reviews, 2008. **88**(2): p. 557-579.
147. Aubert, G., et al., *Collapse of Telomere Homeostasis in Hematopoietic Cells Caused by Heterozygous Mutations in Telomerase Genes.* PLoS Genet, 2012. **8**(5): p. e1002696.
148. Kimura, M., et al., *Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths.* Nat. Protocols, 2010. **5**(9): p. 1596-1607.
149. Oexle, K., *Telomere length distribution and Southern blot analysis.* Journal of theoretical biology, 1998. **190**(4): p. 369-377.
150. Baird, D.M., et al., *Extensive allelic variation and ultrashort telomeres in senescent human cells.* Nat Genet, 2003. **33**(2): p. 203-207.
151. Beh, C.W., et al., *Direct Interrogation of DNA Content Distribution in Nanoparticles by a Novel Microfluidics-Based Single-Particle Analysis.* Nano letters, 2014. **14**(8): p. 4729-4735.
152. Zheng, Y.L., et al., *Less efficient G(2)-M checkpoint is associated with an increased risk of lung cancer in African Americans.* Cancer Research, 2005. **65**(20): p. 9566-9573.
153. Zheng, Y.L., et al., *Telomere deficiencies on chromosomes 9p, 15p, 15q and Xp: potential biomarkers for breast cancer risk.* Human Molecular Genetics, 2011. **20**(2): p. 378-386.
154. Zander, C., R.R. Keller, and J. Enderlein, *Single molecule detection in solution*. Vol. 43. 2002: Wiley Online Library.
155. Friedrich, S.M., K.J. Liu, and T.-H. Wang, *Single Molecule Hydrodynamic Separation Allows Sensitive and Quantitative Analysis of DNA Conformation and Binding Interactions in Free Solution.* Journal of the American Chemical Society, 2015.
156. Turban, L., *Conformation of confined macromolecular chains: cross-over between slit and capillary.* Journal de Physique, 1984. **45**(2): p. 347-353.
157. Tang, J., et al., *Revisiting the Conformation and Dynamics of DNA in Slitlike Confinement.* Macromolecules, 2010. **43**(17): p. 7368-7377.
158. Hong, S., et al., *The Binding Avidity of a Nanoparticle-Based Multivalent Targeted Drug Delivery Platform.* Chemistry & Biology, 2007. **14**(1): p. 107-115.
159. Nasri, M. and D. Thomas, *Relaxation of recognition sequence of specific endonuclease HlndIII.* Nucleic Acids Research, 1986. **14**(2): p. 811-821.
160. Kaul, Z., et al., *Five dysfunctional telomeres predict onset of senescence in human cells.* EMBO reports, 2012. **13**(1): p. 52-59.
161. Pickett, H.A., et al., *Control of telomere length by a trimming mechanism that involves generation of t-circles.* The EMBO Journal, 2009. **28**(7): p. 799-809.
162. Vasan, R.S., et al., *Association of leukocyte telomere length with circulating biomarkers of the renin-angiotensin-aldosterone system: the Framingham Heart Study.* Circulation, 2008. **117**(9): p. 1138-1144.
163. Li, A.Y.-J., et al., *High-mobility group A2 protein modulates hTERT transcription to promote tumorigenesis.* Molecular and cellular biology, 2011. **31**(13): p. 2605-2617.
164. Li, B. and T. De Lange, *Rap1 affects the length and heterogeneity of human telomeres.* Molecular biology of the cell, 2003. **14**(12): p. 5060-5068.
165. Cristofari, G. and J. Lingner, *Telomere length homeostasis requires that telomerase levels are limiting.* The EMBO journal, 2006. **25**(3): p. 565-574.
166. Lai, T.-P., et al., *A method for measuring the distribution of the shortest telomeres in cells and tissues.* Nature communications, 2017. **8**(1): p. 1356.

167. RAY, A. and B. NORDÉN, *Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future.* The FASEB Journal, 2000. **14**(9): p. 1041-1060.
168. Park, H., et al., *Effect of ionic strength on PNA-DNA hybridization on surfaces and in solution.* Biointerphases, 2007. **2**(2): p. 80-88.
169. Tackett, A.J., D.R. Corey, and K.D. Raney, *Non-Watson–Crick interactions between PNA and DNA inhibit the ATPase activity of bacteriophage T4 Dda helicase.* Nucleic Acids Research, 2002. **30**(4): p. 950-957.
170. Allsopp, R.C. and C.B. Harley, *Evidence for a critical telomere length in senescent human fibroblasts.* Experimental cell research, 1995. **219**(1): p. 130-136.
171. Lauzon, W., et al., *Flow cytometric measurement of telomere length.* Cytometry, 2000. **42**(3): p. 159-164.
172. Margolis, R.L. and C.A. Ross, *Diagnosis of Huntington disease.* Clinical chemistry, 2003. **49**(10): p. 1726-1732.
173. Zheng, Y.-L., et al., *Telomere length in blood cells and breast cancer risk: investigations in two case–control studies.* Breast Cancer Research and Treatment, 2010. **120**(3): p. 769-775.
174. Gelfand, Y., A. Rodriguez, and G. Benson, *TRDB—the tandem repeats database.* Nucleic acids research, 2007. **35**(suppl 1): p. D80-D87.
175. Fondon, J.W. and H.R. Garner, *Molecular origins of rapid and continuous morphological evolution.* Proceedings of the National Academy of Sciences, 2004. **101**(52): p. 18058-18063.
176. Dewoody, J., J.D. Nason, and V.D. Hipkins, *Mitigating scoring errors in microsatellite data from wild populations.* Molecular Ecology Notes, 2006. **6**(4): p. 951-957.
177. Zane, L., L. Bargelloni, and T. Patarnello, *Strategies for microsatellite isolation: a review.* Molecular ecology, 2002. **11**(1): p. 1-16.
178. McIver, L., et al., *Population-scale analysis of human microsatellites reveals novel sources of exonic variation.* Gene, 2013. **516**(2): p. 328-334.
179. Duitama, J., et al., *Large-scale analysis of tandem repeat variability in the human genome.* Nucleic acids research, 2014. **42**(9): p. 5728-5741.
180. Beh, C.W., et al., *Fluorescence spectroscopic detection and measurement of single telomere molecules.* Nucleic Acids Research, 2018. **46**(19).
181. Treangen, T.J. and S.L. Salzberg, *Repetitive DNA and next-generation sequencing: computational challenges and solutions.* Nature Reviews Genetics, 2012. **13**(1): p. 36-46.
182. Luke, B. and J. Lingner, *TERRA: telomeric repeat-containing RNA.* Embo Journal, 2009. **28**(17): p. 2503-2510.

Curriculum Vitae

Ye Zhang

Address: 3213 N Calvert St, Baltimore, MD 21218
Cell: (443) 4332383
Email: yzhan119@jhmi.edu

EDUCATION

Johns Hopkins University, School of Medicine	Baltimore, MD
Ph.D. in Biomedical Engineering	Expected March 2019
Concurrent Master of Health Science (MHS) in Biostatistics	Expected May 2019
Tsinghua University	Beijing, China
B.A. in Biomedical Engineering, GPA 92.7/100 (ranked #1 of 30 in department)	July, 2011

RESEARCH EXPERIENCE

Johns Hopkins University, Biomedical Engineering

Graduate Student with Prof. Tza-Huei Wang in BioMEMS & Single Molecule Dynamics Lab

- Application of functional nanoparticles, microfluidics and single molecule confocal spectroscopy for multiplex nucleic acid analyses and molecular study of diseases
- Developed a novel silica nanomembrane material providing greatly enhanced extraction yields of highly-pure, high-molecular weight DNA
- Employed a quantum-dot based FRET-fluorescence scheme for multiplex genetic and epigenetic assays
- Established a novel method of fluorescence encoding for multiplexed padlock probe diagnostic assays
- Leveraged confocal single-molecule fluorescence spectroscopy for absolutely quantitative assessment of telomere length
- Fabricated digital microfluidic devices to achieve highly-multiplexed detection of infectious diseases and epigenetic biomarkers in ovarian cancer

Tsinghua University, Biomedical Engineering

Undergraduate Student in Prof. Jing Liu's lab and Prof. Jing Cheng's lab

- Employed functional nanoparticles for advanced physical hyperthermia treatments of cancer
- Developed a microfluidic system with an isothermal detection scheme for real-time detection of cancer biomarkers

PUBLICATIONS (*co-first author)

1. **Y. Zhang***, P. Zhang* and T. -H. Wang. Droplet-Based Digital Ratiometric Fluorescence Coding for Multiplex Nucleic Acid Amplification Testing (in preparation).
2. **Y. Zhang**, L. Chen, K. Hsieh and T. -H. Wang. Ratiometric Fluorescence Coding for Multiplex Nucleic Acid Amplification Testing. Analytical Chemistry, 2018; 90(20): 12180-12186.
3. C. W. Beh*, **Y. Zhang***, Y. -L. Zheng, B. Sun, and T. -H. Wang. Fluorescence Spectroscopic Detection and Measurement of Single Telomere Molecules. Nucleic Acid Research, 2018; gky627.
4. **Y, Zhang***, Y. Zhang*, J. M. Burke, K. Gleitsman, S. M. Friedrich, K. J. Liu, and T. -H. Wang. A Simple Thermoplastic Substrate Containing Hierarchical Silica Lamellae for High-Molecular-Weight DNA Extraction. Advanced Materials, 2016; 28(48): 10630-36.
5. T. R. Pisanic II, **Y. Zhang** and T. -H. Wang. Quantum dots in diagnostics and detection: principles and paradigms. Analyst, 2014; 139(12): 2968-81.
6. F. Tang, **Y. Zhang**, J. Zhang, JW. Guo and R. Liu. Assessment of the efficacy of laser hyperthermia and nanoparticle-enhanced therapies by heat shock protein analysis. AIP Advances, 2014; 4(3).
7. L. Zhang, Y. Song, T. Fujita, **Y. Zhang**, M. Chen, T.-H. Wang. Large Enhancement of Quantum Dot Fluorescence by Highly Scalable Nanoporous Gold. Advanced Materials, 2014; 26(8): 1289-94
8. TZ. Wang, **Y. Zhang**, C. Wang, L. Xie, L. Ma, ZY. Li, XB. Luo, H. Tian, Q. Li, X. Li, ZY. Lv, XF. Bao and GL. Huang Detect early stage lung cancer by a LAMP microfluidic chip system with a real-time fluorescent filter processor. Science China Chemistry, 2012; 55(4): 508-14.

CONFERENCE PROCEEDINGS AND PRESENTATIONS

Peer-reviewed Conference Publications

1. **Y. Zhang**, P. Zhang and T. H. Wang. "Droplet-based digital ratiometric fluorescence coding for multiplex nucleic acid amplification testing. 32nd International Conference on Micro Electro Mechanical Systems (IEEE MEMS 2019), 2019
2. C.W. Beh, **Y. Zhang**, Y.L. Zheng, B. Sun, T.H. Wang, "Fluorescence spectroscopic detection and measurement of single telomere molecules", 13th IEEE International Conference on Nano/Micro Engineered and Molecular Systems (IEEE NEMS 2018), p. 28-31, 2018
3. C.W. Beh, **Y. Zhang**, T.H. Wang, "Fluorescence Flow Moriometry – Flow Cytometry-Like Analysis Method for Single Molecule and Particle Cauterization", 20th International Conference on Miniaturized Chemical and Biochemical Analysis Systems (micro-TAS 2016), p. 1362-1363, 2016

4. Y. Zhang, **Y. Zhang**, B. Keeley, A. Stark and T.H. Wang, “Spontaneous Shrinking Silica Nanomembrane for Solid Phase”, 8th Annual IEEE International Conference on Nano/Micro Engineered and Molecular Systems (IEEE NEMS 2013), p. 444-445, 2013
5. Y. Zhang, **Y Zhang**, T.H. Wang, “Hierarchical Silica Nanomembrane Driven by Thermal Shrinkage and its Application for Solid Phase DNA Extraction”, Proc. 13th IEEE International Conference on Nanotechnology (IEEE NANO 2013), 2013

Conference Abstracts and Presentations

1. Y, Zhang, **Y. Zhang**, J. M. Burke, K. Gleitsman, S. M. Friedrich, K. J. Liu, and T. -H. Wang. Shrinking-Induced Silica Lamella for High-Molecular Weight DNA Extraction. IEEE Nanotechnology Materials and Devices Conference (NMDC), 2017, (October 2nd, 2017).
2. **Y. Zhang**, C. W. Beh, K. Liu and T. –H. Wang. Highly Quantitative Cylindrical Illumination Confocal Spectroscopy (CICS) for Single Molecule Analysis of Genomic Content. Chesapeake Bay Area Single Molecule Biology Meeting, 2017, (April 8th, 2017).
3. **Y, Zhang**, Y. Zhang, K. J. Liu, and T. -H. Wang. Heat-Shrunken Hierarchical Silica Nanomembrane for Solid Phase DNA Extraction. NanoEngineering for Medicine and Biology Conference (NEMB), 2016, (February 26th, 2016).
4. **Y, Zhang**, Y. Zhang, K. J. Liu, and T.-H. Wang. Heat-Shrunken Hierarchical Silica Nanomembrane for Solid Phase DNA Extraction. Biomedical Engineering Society (BMES), 2014, (October 22nd, 2014).
5. **Y. Zhang**, Y. Zhang, and T. -H. Wang. Hierarchical Nanomembrane Driving by Heat-shrinkage of Polyolefin Film and Application for Solid Phase DNA Isolation. Mid-Atlantic Micro/Nano Alliance Symposium (MAMNA), 2014, (March 3rd, 2014)

AWARDS

- Materials Research Society (MRS) Spring Meeting Science as Art Finalist, 2017
- China National Scholarship (highest scholarship in China; top 2% of all university students), 2011
- Third Prize in Mathematical Contest in Modeling (MCM) & Interdisciplinary Contest in Modeling (ICM), 2010
- Goldman Sachs Global Leaders Program Scholarship, awarded to only 150 of all the sophomores all over the world per year, 2009
- Bronze Medal of the Chinese Mathematic Olympics, 2007

TEACHING AND MENTORING EXPERIENCE

Johns Hopkins University, Biomedical Engineering Baltimore, MD

- ***Teaching Assistant*** 2014 -2016

Taught, evaluated and graded undergraduate students as a teaching assistant for three Bioengineering courses

- ***Mentor, Women in Science and Engineering (WISE) Program*** 2017 -2018

  Supervised a high school student in microfluidic device fabrication and PCR practice two afternoons per week

- ***Mentor, Study Buddy Tutoring Program at Hampden Family Center*** 2016

  One-on-one tutored a 4th grade student in his mathematics and helped him with his homework

- ***Mentor, STEM Achievement in Baltimore Elementary School (SABES)*** 2014

  Guided 3rd grade students to understand and improve interests in STEM through fun science experiments and inquiry-led projects relevant to their community